The Dissertation Committee for Yen-I Grace Chen Certifies that this is the approved version of the following dissertation:

Proteomic Analysis of the Pre-mRNA Splicing Machinery Utilizing Chromosomal Locus Epitope Tagging in Metazoans

Committee:

Scott Stevens, Supervisor

Henry Bose

Robert Krug

Arlen Johnson

David Hoffman

Proteomic Analysis of the Pre-mRNA Splicing Machinery Utilizing Chromosomal Locus Epitope Tagging in Metazoans

by

Yen-I Grace Chen, M.D.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

August, 2007

Dedication

To my loving family

Acknowledgements

I offer sincere thanks to my advisor, Professor Scott Stevens for providing me the opportunity to join his laboratory. His enthusiasm and patience contributed to the success of my research. I also wish to acknowledge professors Robert Krug, Henry Bose, Arlen Johnson, and David Hoffman for serving as my dissertation committee and for their constructive comments and expertise during my dissertation work; professors Phil Tucker and Karen Artzt for helpful conversations, and William Kuziel for plasmids.

I would especially thank Shanna Maika and Deb Surman for their persistence and painstaking work in generating our transgenic mice, breeding, and maintenance; Xin Yao and Chad McKee for technical assistance. I also wish to thank all my colleagues in the Stevens lab: Josh Combs, Adam Roth, Champ Gupton, Rea Lardelli, Jennifer Hennigan and former members Dannialle Clayton and Colin Chu, for helpful discussions and establishing a fun environment for research.

I would also like to thank Roger Moore, Helen Ge, and Terry Lee for mass spectrometry expertise and Jimi Lynn Brandon for mouse tissue processing. My gratitude also extends to all members of Bose lab for chicken DT40 cell advice and tissue culture help and all members of Johnson lab for their suggestions and friendship over the years.

Lastly, I would like to thank my dear husband, Edward, for his encouragement and love during my graduate studies; my parents for their support and belief in my achievement and my sister, for sharing laughs and tears with me in Austin.

Proteomic Analysis of the Pre-mRNA Splicing Machinery Utilizing Chromosomal Locus Epitope Tagging in Metazoans

Publication No._____

Yen-I Grace Chen, Ph.D.

The University of Texas at Austin, 2007

Supervisor: Scott W. Stevens

Epitope tagging in metazoans is an important tool for biochemical analyses and is generally accomplished by using trans-genes with in-frame epitope tags. However, protein levels from trans-genes are rarely representative of native levels. To overcome the shortcomings using trans-genes, epitope tags were introduced by homologous recombination technology, termed CLEP tagging (Chromosomal Locus EPitope tagging), immediately upstream of the stop codon of targeted genes in chicken B cell line DT40 and mouse embryonic stem (ES) cells. I first demonstrated the feasibility and promise of this technique in DT40 cells by purifying low abundance polypeptides and factors loosely associated with the SmD3 protein, a core protein participating in pre-mRNA splicing and mRNA turnover, with a TAP (tandem affinity purification) tag. Glycerol gradient separation was performed to further characterize the SmD3-associated protein complexes from the 200S fractions, corresponding to the supraspliceosomes. The purification included all five spliceosomal snRNAs. Most known snRNP-associated

proteins, 5’ end binding factors, 3’ end processing factors, mRNA export factors, hnRNPs, and other RNA binding proteins were identified from the protein components. Intriguingly, the purified supraspliceosomes also contained a number of structural proteins, nucleoporins, chromatin remodeling factors, and several novel proteins that were absent from splicing complexes assembled *in vitro*. I showed that the *in vivo* analyses provide a more comprehensive list of polypeptides associated with pre-mRNA splicing apparatus as well as those that coupled transcription to the pre-mRNA processing steps. With similar techniques, the TAP tag was inserted into the chromosomal locus of a pre-mRNA splicing factor component, mSART-1 in live mice. Surprisingly, a profound autoimmune response was induced in homozygous-modified mice, due likely to an inappropriate stimulation of the immune system. I believe these mice will serve as a valuable tool for the studies of mammalian autoimmune diseases, especially those resulting from the generation of autoantibodies against RNP components.

Table of Contents

List of Tables

List of Figures

List of Illustrations

Chapter 1: Introduction

1.1 GENE EXPRESSION IN EUKARYOTES

1.1.1 Overview

Gene expression in eukaryotes is a highly dynamic and regulated process. In order to have proper gene expression in eukaryotic cells, RNA polymerase II (RNA Pol II) transcripts undergo several coordinated processing steps before the mature mRNA is exported to the cytoplasm for translation. These pre-mRNA processing steps include capping of the pre-mRNA with a 7-methyl guanosine cap structure at the 5' end, intron removal by the spliceosome, RNA editing, and cleavage and polyadenylation of pre-mRNA by the 3' end processing machinery. Studies have shown that these steps are coupled to each other and occur co-transcriptionally on the nascent transcripts [Illustration 1.1, for review, see (Howe, 2002; Maniatis and Reed, 2002; McCracken et al., 1997)].

1.1.2 The carboxyl-terminal domain (CTD) of RNA Pol II: an assembly platform for pre-mRNA processing factors

The carboxyl-terminal domain (CTD) of the RNA Pol II plays an important role for the recruitment and regulation of processing factors. The CTD is composed of heptad consensus repeats, YSPTSPS, which is repeated 27 times in yeast and 52 times in humans. Serine 2 (Ser2) and Serine 5 (Ser5) residues within the repeats are phosphorylated during transcription. The pattern of the CTD phosphorylation changes as transcription proceeds (Komarnitsky et al., 2000). Ser5 phosphorylation is detected primarily at the promoter while Ser2 phosphorylation is seen only in coding regions. In

yeast, the recruitment of Kin28 kinase, a subunit of transcription factor TFIIH, phosphorylates Ser5 of the CTD at the promoter (Hengartner et al., 1998) and as elongation proceeds, phosphorylation of the CTD on Ser2 by Ctk1 increases (Cho et al., 2001). In higher eukaryotes, several C-terminal kinases (CTK) have been found to catalyze phosphorylation and their recruitment and affinity are affected by the position of the targeted repeats within the CTD and the phosphorylation state of the neighboring residues (Palancade and Bensaude, 2003; Ramanathan et al., 2001). The CTD deletions do not inactivate transcription but prevent efficient co-transcriptional capping, splicing and 3' end cleavage and polyadenylation (Fong and Bentley, 2001). Moreover, dynamic site-specific phosphorylation and dephosphorylation of the CTD is critical for coupling RNA Pol II transcription to the processing factors (Maniatis and Reed, 2002).

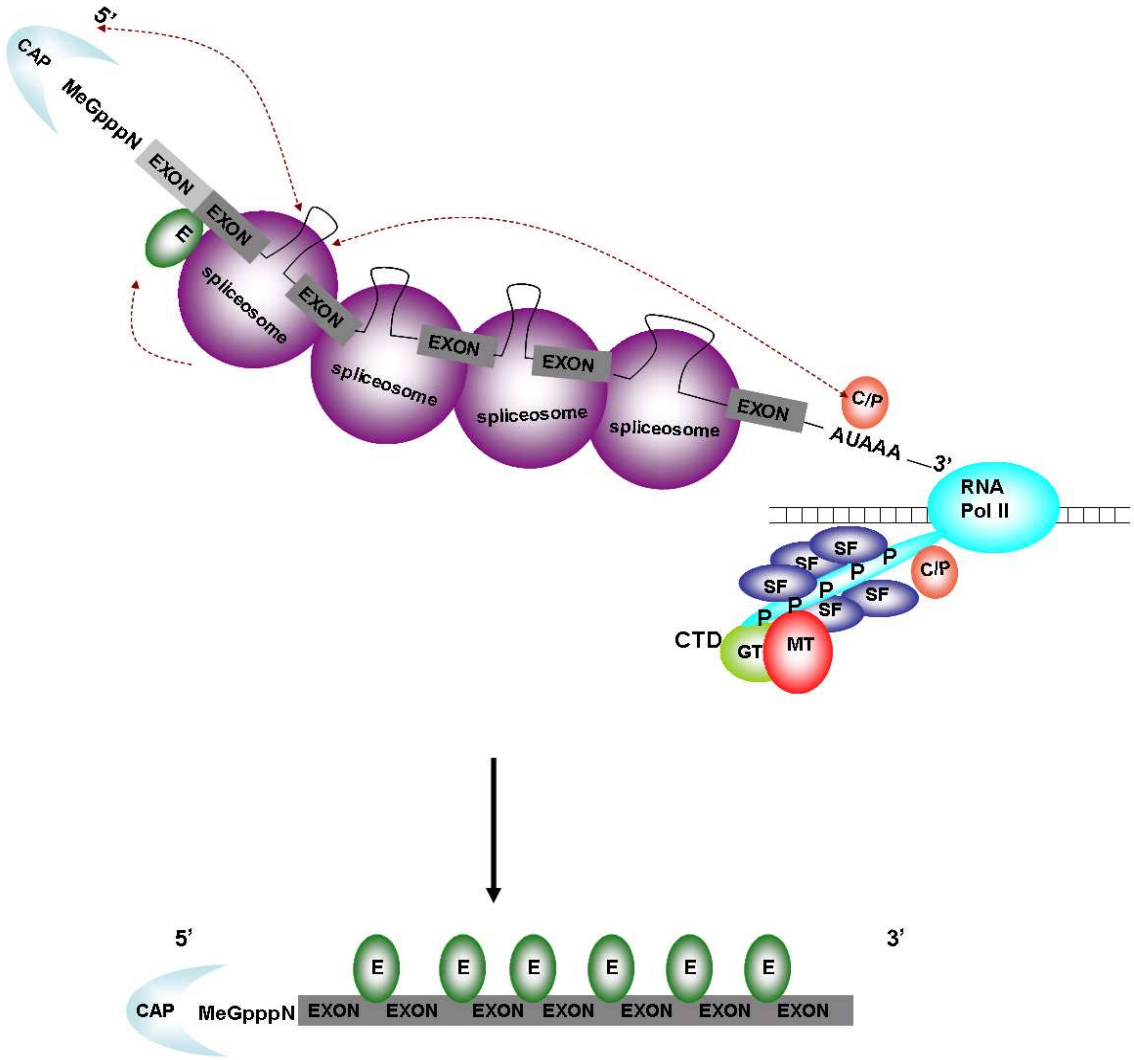

Illustration 1.1 Coupled interaction in gene expression

The C-terminal domain (CTD) functions as an assembly platform and regulator for co-transcriptional recruitment of pre-mRNA processing factors. The 5' capping enzymes, guanylyltransferase (GT) and 7-methyltransferase (MT), splicing factors (SF) and 3' end cleavage/polyadenylation factors (C/P) are recruited via the CTD. Dash arrows indicate that all the processing pathways are coupled with each other for proper gene expression. Exon-Exon junction complexes (E) formed during splicing are important for mRNA export.

1.1.3 Coupling between transcription and the pre-mRNA processing machinery

1.1.3.1 5' end capping

Capping of the nascent transcripts at the 5' end protects them from degradation. This step involves 3 enzymes, the guanylyltransferase, the 7-methyltransferase, and the RNA triphosphatase. Different phosphorylated/dephosphorylated forms of the CTD regulate the recruitment of the capping enzymes. Ser5 phosphorylation on the CTD is necessary for the recruitment of the capping enzymes, guanylyltransferase and 7-methyltransferase. Subsequent Ser5 dephosphorylation in early elongation results in the release of the guanylyltransferase (Komarnitsky et al., 2000; Schroeder et al., 2000). Ser2 on the CTD is then phosphorylated, leading to the recruitment of the RNA processing factors required for the subsequent steps (Cho et al., 2001). Capping enzymes also cooperate with transcription factors to stimulate elongation and promoter clearance (Mandal et al., 2004).

1.1.3.2 Recruitment of the spliceosome

Several splicing factors have been shown to be coupled to the transcription machinery. The U1 snRNP associates with transcription elongation factors at the time of recognition of the 5' splice site of the intron (Kotovic et al., 2003). The U1 snRNP-associated splicing factor Prp40 (Kao and Siliciano, 1996), which is proposed to bring the 5' and the 3' splice sties together during splicing (Berglund et al., 1997), interacts with phosphorylated CTD (Morris and Greenleaf, 2000). Elongation factor, TAT-SF1 and CUS2 (yeast homolog), interacts with splicing factors and regulates the assembly of splicing complexes (Fong and Zhou, 2001; Yan et al., 1998). SR proteins, important for both constitutive and alternative splicing (Graveley, 2000), have been shown to interact with the transcriptional coactivator PGC-1 (Cramer et al., 1999), suggesting a connection

between transcription and alternative splicing (Goldstrohm et al., 2001). In addition, the cap-binding complex has been shown to interact with splicing factors and promote the recognition of the proximal 5' splice site (Colot et al., 1996; Lewis et al., 1996). Coupling of splicing factors to transcription is proposed to concentrate splicing factors on nascent transcripts, allowing constitutive and alternative splicing to occur efficiently and accurately (Maniatis and Reed, 2002).

1.1.3.3 3' end formation

The formation of the 3' end of mRNA occurs in two steps. First, the nascent transcript is cleaved 20-30 bases downstream of a conserved poly(A) site, AAUAAA, by the cleavage and polyadenylation factors. Second, poly(A) polymerase adds a poly(A) tail to the 3' OH of the nucleotide at the site of mRNA cleavage. It has been shown that 3' end processing factors, CstF50, Ydh1, Yhh1, and Rna14 bind to the CTD (Proudfoot, 2004). Splicing and 3' end formation have also been shown to be coupled (Vagner et al., 2000). Splicing factors are able to interact with poly(A) polymerase and promote 3'end formation; polyadenylation, in turn, can stimulate splicing [reviewed in (Berget, 1995)].

1.1.3.4 Recruitment of mRNA export factors

After transcription termination, the mature mRNA and RNA Poll II are released from the DNA template. The mature RNA is then transported to nuclear pores for export to the cytoplasm. Evidence has shown that mRNAs generated by splicing are more efficiently exported than their intronless counterparts transcribed from DNA, even when they are identical in sequence (Luo and Reed, 1999). Protein complexes recruited adjacent to the exon-exon junctions during splicing, are assembled into a unique complex and critical to couple splicing to mRNA export (Le Hir et al., 2000). UAP65, a splicing

factor required for early spliceosome assembly, is also a protein component in the exon junction complex that couples splicing to mRNA export (Luo et al., 2001; Zhou et al., 2000). UAP65 interacts directly with the mRNA transport factor, ALY, which then interacts with the nuclear export adapter TAP, delivering the mRNA to the nuclear pore for export (Strasser and Hurt, 2001). Links between transcription, 5' end capping, splicing, 3' end formation and export ensure all the steps are timely coordinated for proper gene expression.

1.2 PRE-MRNA SPLICING AND THE SPLICEOSOME ASSEMBLY

1.2.1 Small nuclear RNAs (snRNA)

1.2.1.1 The discovery of snRNAs

The existence of low molecular weight RNA, 90-400 nucleotides long, which sedimented in the 4-8S region of a sucrose gradient was reported in the 1960s (Hodnett and Busch, 1968; Moriyama et al., 1969; Muramatsu et al., 1966; Prestayko and Busch, 1968). These small RNAs were divided into small nucleolar, small nuclear and small cytoplasmic RNAs based on their cellular localization. A subset of these small RNAs, rich in uridylic acid was first identified from nucleoli of Novikoff hepatoma cells (Muramatsu et al., 1966). This RNA had an uridine-rich base composition, different from that of the ribosomal RNA (rRNA) and transfer RNA (tRNA) and was designated U-RNA. By sedimentation, chromatography, and terminal nucleoside analysis, six U-RNA (U-small nuclear RNA, U-snRNA), U1 snRNA (Reddy et al., 1981b; Reddy et al., 1974), U2 snRNA (Shibata et al., 1975), U3 snRNA (Reddy et al., 1979; Reddy et al., 1980), U4 snRNA (Reddy et al., 1981a), U5 snRNA (Krol et al., 1981), and U6 snRNA (Epstein et al., 1980) were identified and well characterized from nuclei. U3 is

nucleolar while U1, U2, U4, U5, and U6 are nucleoplasmic. snRNAs are present at 0.2-1× 10^6 molecules per cell. The sequencing of snRNAs led to the discovery of cap structures (Reddy et al., 1974), followed by studies on cap structure of messenger RNA (mRNA) and viral RNA (Rottman et al., 1974; Shatkin, 1976).

1.2.1.2 The biogenesis of snRNAs

With the exception of U6, the U1, U2, U3, U4, and U5 are synthesized by RNA polymerase II, acquiring a monomethyl guanosine (7mGpppG) cap and their primary transcripts are extended far beyond the 3' end of the mature RNA (Busch et al., 1982; Eliceiri, 1980). They are compacted into a large export complex, including cap-binding proteins and CRM1/RanGTP for export to the cytoplasm (Kiss, 2004). In the cytoplasm, the snRNAs associate with the survival of motor neuron (SMN) complex (Paushkin et al., 2002), which directs the assembly of the seven related Sm proteins SmB/B', SmD1, SmD2, SmD3, SmE, SmF, and SmG, which assemble into a heteroheptameric ring complex and bind to the conserved Sm-binding site within the snRNA (Luhrmann et al., 1990; Will and Luhrmann, 2001). Binding of the Sm proteins is important for hypermethylation of the 7mG cap to 2,2,7-trimethylguanosine (TMG) cap (Mattaj, 1986; Mouaikel et al., 2002; Plessel et al., 1994), as well as exonucleolytic cleavage of 3' sequences. Both the assembled Sm core and TMG provide signals for the efficient import of the newly assembled small nuclear ribonucleoproteins (snRNPs). The snRNPs are then imported back to the nucleoplasm with the help of snurportin and importin β (Palacios et al., 1997) for further modification in Cajal bodies (Jady et al., 2003). U6, synthesized by RNA polymerase III, carries a γ–monomethylphosphate (mpppG) cap structure (Singh and Reddy, 1989). A heteroheptameric ring of seven Sm-like proteins, Lsm2, Lsm3, Lsm4, Lsm5, Lsm6, Lsm7 and Lsm8, binds to the 3'-terminal

UUUU sequence of U6 snRNA (Achsel et al., 1999). Binding of the Lsm proteins is critical to direct the modification of U6 in the nucleolus (Gerbi et al., 2003). The biogenesis of U6 is confined to the nucleus (Gerbi et al., 2003; Kiss, 2004).

1.2.2 Small nuclear ribonucleoproteins (snRNP)

1.2.2.1 The isolation of snRNPs

The snRNAs are not present as naked RNA molecules in the nucleoplasm but exist in specific RNA-protein complexes called small nuclear ribonucleoprotein (snRNP) particles (Lerner and Steitz, 1979; Raj et al., 1975). Isolation of the U1 and U2 snRNPs was first achieved by Raj *et al.* using Sepharose gel filtration and sucrose density gradient and 10 proteins were shown to be present in the complexes (Raj et al., 1975). The discovery of an autoimmune reaction between several classes of eukaryotic snRNPs and autoantibodies from sera of patients with rheumatic diseases such as systemic lupus erythematosus, mixed connective tissue diseases, scleroderma, and Sjögren's syndrome has facilitated the structure and localization analyses of snRNP particles. Lerner and Steitz showed that anti-Sm antibodies immunopurified the U1, U2, U4, U5, and U6 snRNP particles whereas anti-U1 RNP antibodies precipitated only U1 snRNP (Lerner and Steitz, 1979). The protein components are essential for antigenicity and antigenic determinants of the snRNP are highly conserved between species. Another two classes of small ribonucleoproteins, Ro/SSA and La/SSB are also detected by antibodies associated with Sjögren's syndrome and lupus erythematosus (Lerner et al., 1981). Immunological methods offer a useful approach for snRNP purification; however, it is difficult to recover snRNPs in a native structure from the antibody affinity columns.

In order to analyze the native structure and function of snRNPs, The Lührmann laboratory developed a procedure for the isolation of snRNPs on a preparative scale using

antibodies against the TMG cap, anti-$m_3^{2,2,7}$ G antibodies, and purified proteins by affinity chromatography with anti-$m_3^{2,2,7}$ G IgG (Bringmann et al., 1983). Eluting the antibody-bound snRNPs from the column with excess nucleoside $m_3^{2,2,7}$ G or 7mG maintains the native structure of the snRNPs.

1.2.2.2 A link between snRNPs and splicing

The first study to implicate the role of snRNPs in pre-mRNA processing pathways was reported by Sekeris *et al.* who showed that the snRNAs co-sediment with large particles that bind hnRNA (Sekeris and Niessing, 1975). In the early 1980s, Lerner *et al.* reported that snRNAs were present in all eukaryotes and are well conserved through evolution. snRNPs were present in highest abundance in metabolically active cell types, such as liver cells and most importantly, the 5' end sequence of the U1 snRNA is complementary to 5' splice junctions in mRNA precursors (Lerner et al., 1980). This was the first demonstration that an RNA-RNA interaction is important for molecular recognition. Another clue to the role of the snRNPs in splicing came from the evidence that splicing was inhibited in HeLa cell nuclei when they are preincubated with anti-Sm or anti-RNP antibodies (Yang et al., 1981). Further studies of the structure and function of snRNPs have revealed different roles of distinct snRNPs in the pre-mRNA splicing reaction as described in the text below.

1.2.3 The splicing reaction

1.2.3.1 Pre-mRNA splicing machinery, spliceosome

The coding sequences, exons, in most primary transcripts of RNA polymerase II are interrupted by intervening sequences, IVSs, or introns. Introns have to be removed from pre-mRNA in a process called pre-mRNA splicing to generate the mature mRNA

before it is exported from the nucleus to the cytoplasm where it is translated into protein (Sharp, 1994). The first evidence revealing that coding regions were interrupted was found by the Sharp and the Roberts laboratories in 1977 (Berget et al., 1977; Chow et al., 1977). They showed the DNA template of adenovirus major late was not able to form a smooth hybrid with its mRNA, indicating that some portions of the sequence, later termed introns, were missing from the mRNA. In addition, three adenovirus major late mRNAs shared a common 5' end.

The machinery that splices pre-mRNA is called the spliceosome which is comprised of hundreds of proteins and five small nuclear RNAs (U1, U2, U4, U5, and U6) assembled into small nuclear ribonucleoproteins (snRNPs). U1, U2, U4, and U5 are transcribed by RNA polymerase II, sharing two common features, the 5'-terminial TMG cap and highly conserved Sm site [RA$(U)_4$GR] flanked by two stem-loop structures (Luhrmann et al., 1990). U6 is transcribed by RNA polymerase III, containing a γ-mono-methyl-phosphate cap and no Sm site. Sm proteins are common proteins in U1, U2, U4, U5 snRNPs and are essential for 5' hypermethylation of the 7mG cap to 2, 2, 7-trimethylguanosine and re-import of the mature snRNAs back to the nucleus. Other proteins are found associated with specific snRNPs, involved in different dynamic interaction with pre-mRNA.

1.2.3.2 Splicing signals

The splicing signals are consensus sequences located at the exon/intron junctions [(Stephens and Schneider, 1992) and Illustration 1.2].

5'-AG/GURAGU-------intron------YNCURAY-(Y)n---------YAG/-3'

The 5' splice site conforms to the consensus sequence AG/GURAGU where the slashes are the exon/intron borders, R is any purine, and Y is any pyrimidine. The 3'

splice site is characterized by the sequence YAG. The first two and the last two bases, underlined GU and AG, respectively, of the intron are almost invariant. In addition to the consensus sequences at the 5' and 3' end of the introns, another two sequence elements, the polypyrimidine tract [(Y)n] and branchpoint sequences (YNCURAY) are also important for accurate splicing (Kramer, 1996). The polypyrimidine tract is a stretch of pyrimidine residues preceding the 3' splice site (Kramer, 1996). It is important for the definition of the 3' splice site and the recruitment of proteins for the earliest stage of spliceosome assembly. The branchpoint sequence YNCURAY is located 18-40 nucleotides upstream of the 3' splice site and the branchpoint adenine nucleotide, the site of branch formation, is underlined (Zhuang et al., 1989). In the yeast *Saccharomyces cerevisiae*, this sequence is more highly conserved: and is almost always UACUAAC whereas the sequence is more variable in higher eukaryotes. All of these consensus sequences are important for proper splicing and gene expression.

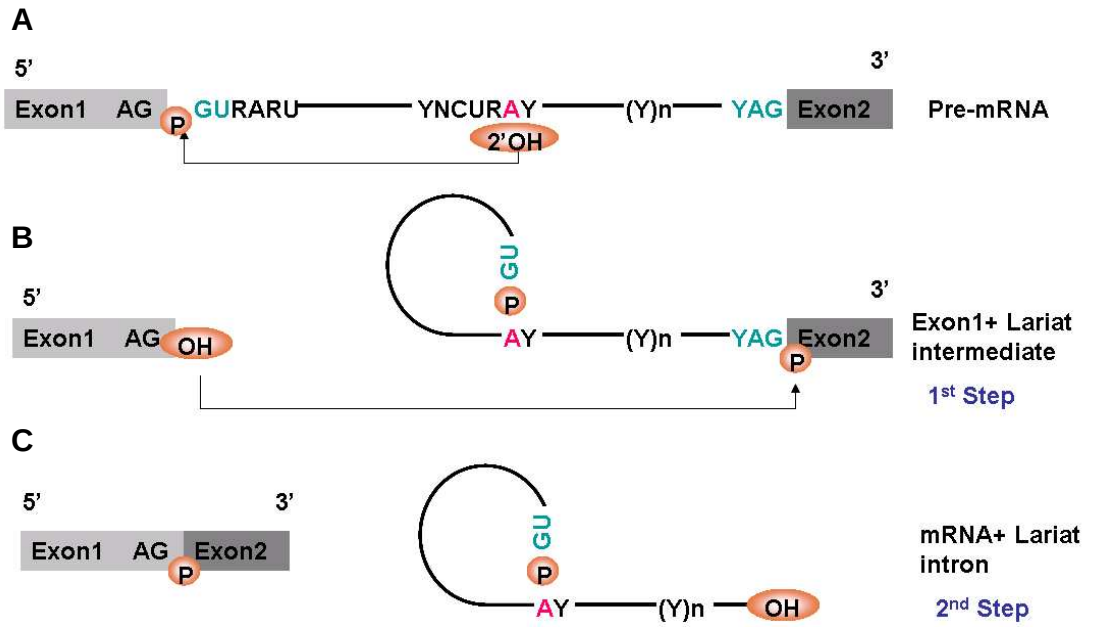

Illustration 1.2 Splicing signals and the two catalytic steps of pre-mRNA splicing

(A) The pre-mRNA is represented by two boxes, exon 1 and exon 2 connected by a line, intron sequences. The conserved splicing signals within the 5' splice site [GU], 3' splice site [AG], branchpoint [YNCURAY], and pyrimidine tract [(Y)n] are highlighted. The arrows represent the nucleophilic attack of the hydroxyl group (OH) on the splice junction. (B) The first step of splicing results in the 5' splice site cleavage and the lariat intermediate formation. (C) The second step results in the 3' splice site cleavage, forming a lariat intron and concurrent exon ligation.

1.2.3.3 The catalysis of splicing

Intron excision involves 2 transesterification steps. The first step of the reaction is initiated by a nucleophilic attack of the 2' hydroxyl group of the adenosine at the branch site on the phosphodiester bond at the 5' splice site, yielding the 5' exon and the lariat intermediate. The lariat intermediate is formed by the guanosine at the 5' terminus of the intron covalently linked to the branch site adenosine in an unusual 2'-5'-

phosphodiester bond. In the second step, the free 3'-hydroxyl group on the 5' exon attacks the phosphodiester bond between the intron and the 3' exon, resulting in the concurrent ligation of the exons and the release of a lariat-shaped intron (Padgett et al., 1984; Ruskin et al., 1984).

1.2.3.4 The spliceosome assembly cycle

Studies have showed that the spliceosome is assembled in a stepwise fashion (Illustration 1.3). First, the U1 snRNP interacts with the 5' splice site via base pairing of the 5' end of U1 snRNA to the conserved sequences at the 5' splice site. The interaction between U1 snRNP and pre-mRNA forms the early (E) complex (Ruby and Abelson, 1988; Seraphin et al., 1988; Zhang and Rosbash, 1999; Zhuang and Weiner, 1986). This stage occurs in an ATP- and temperature-independent manner. Some protein components are recruited in the E complex for the initial recognition event. For example, SR proteins are incorporated into the spliceosome at the time of E complex formation and promote the binding of U1 snRNP to 5' splice sites and splice site selection (Jamison et al., 1995; Kohtz et al., 1994; Zahler and Roth, 1995). U2 snRNP auxiliary factor (U2AF), composed of two subunits of 35 and 65 kDa, binds to the polypyrimidine tract (Zamore and Green, 1989). It is not only important for U1 snRNP recruitment to the 5' splice site but is also essential for the association of between U2 snRNP and the pre-mRNA (Forch et al., 2003; Ruskin et al., 1988; Zamore and Green, 1989). Splicing factor 1 (SF1) specifically recognizes the branch point sequence in the pre-mRNA facilitating the association between U2 snRNP and the pre-mRNA (Berglund et al., 1997; Liu et al., 2001). Stable interaction between the U2 snRNP and the branchpoint sequence forms of the pre-mRNA complex A (Parker et al., 1987; Zhuang and Weiner, 1989). This step, as well as many subsequent steps, is ATP-dependent

(Cheng and Abelson, 1987; Pikielny and Rosbash, 1986). The U4/U6•U5 tri-snRNP then joins in the B complex and forms a mature spliceosome, catalyzing two steps of transesterification for intron excision in the C complex (Kramer, 1996; Padgett et al., 1984; Ruskin et al., 1984).

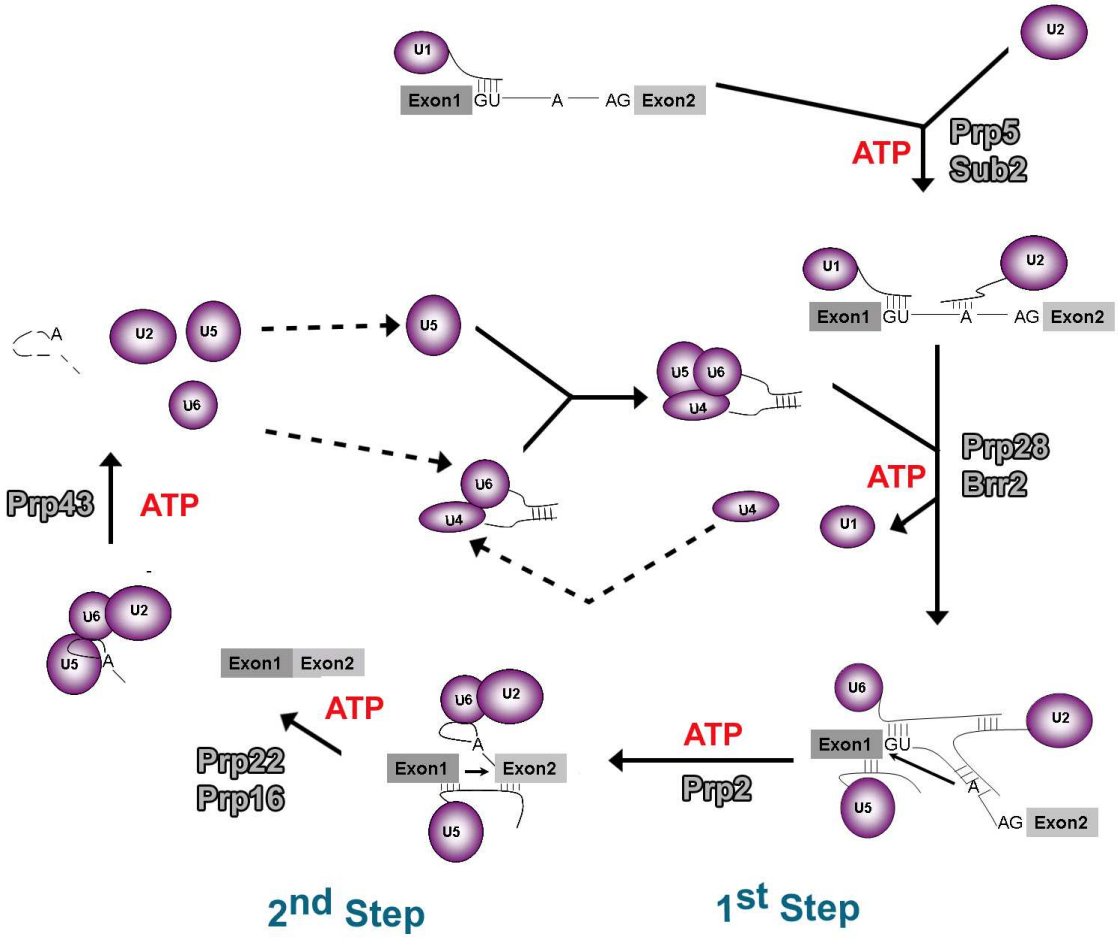

Illustration 1.3 The classical pre-mRNA spliceosome assembly cycle

The pre-mRNA is represented by two boxes, exon 1 and exon 2 connected by a line, intron sequences. The conserved sequences are highlighted and the U1, U2, U5, U4/U6, U4/U6•U5 snRNPs are represented as purple spheres. Base pairing interactions between the pre-mRNA and the snRNAs are indicated by the hatched markings. The ATP and 8 helicases important for RNA arrangement are denoted.

1.2.3.5 Structure and composition of snRNPs

The five spliceosomal snRNPs (U1, U2, U4, U5, and U6), each containing a small stable RNA bound by several proteins, plus numerous other less stably-associated splicing factors are involved in pre-mRNA splicing reaction. U4 and U6 snRNA are extensively base-paired and package into a single ribonucleoprotein particle while U1, U2, and U5 snRNPs can be found in distinct particles (Maniatis and Reed, 1987). With improved isolation techniques, greater numbers of polypeptides have been identified as additional particle-specific proteins.

The U1 snRNP, which sedimented as a 10S particle, was isolated independently from the laboratories of Steitz and Lührmann (Bringmann and Luhrmann, 1986; Bringmann et al., 1983; Hinterberger et al., 1983). It contains the common Sm proteins, SmB', SmB, SmD, SmD', SmE, SmF, SmG and additional three unique polypeptides, U1-70K, U1-A, and U1-C.

The 17S U2 snRNP was identified from Lührmann's laboratory containing the common Sm proteins, and the U2A', U2B'', SF3a, and SF3b components (Behrens et al., 1993a; Behrens et al., 1993b). The integrity of the U2 snRNP appears to be very sensitive to salt concentration as it was first isolated as 12S particle without SF3a and SF3b components (Bringmann and Luhrmann, 1986).

The 25S U4/U6•U5 tri-snRNP particle joins the spliceosome as a tripartite particle (Behrens and Luhrmann, 1991). The tri-snRNP contains additional polypeptides, U4/U6•U5-110K, U4/U6•U5-65K, U4/U6•U5-61K, and U4/U6•U5-15.5K, that are not associated with the 20S U5 (U5-220K, U5-200K, U5-116K, U5-102K, U5-100K, U5-40K and U5-15K) (Bach et al., 1989) or 10S U4/U6 snRNPs (U4/U6-90K and U4/U6-60K). The isolated U5 and U4/U6 RNPs themselves can not assemble into the 25S particle without the help of additional polypeptides released from the tri-snRNP by

micrococcal nuclease treatment. Moreover, addition of U5 and U4/U6 snRNPs cannot restore the splicing activity in extracts devoid of 25S tri-snRNP particle (Utans et al., 1992).

1.2.3.6 ATP-dependent RNA helicases

Temporal and spatial regulation of RNA and protein conformational rearrangements in the pre-mRNA splicing machinery requires energy in the form of ATP at multiple steps (Madhani and Guthrie, 1994). Eight RNA helicase in the yeast *S. cerevisiae*, Prp5p, Sub2p, Prp28p, Brr2p, Prp2p, Prp16p, Prp22p, and Prp43p, containing sequence motifs of the DEAD/H family have been well characterized (Illustration 1.3). Prp5p contains distinct U1- and U2-interacting domains that are required for U2 snRNP association (Dalbadie-McFarland and Abelson, 1990; Xu et al., 2004), a reaction coordinated by Sub2p (Libri et al., 2001). During catalytic activity, Prp28p destabilizes the interaction between U1 snRNA and the 5' splice site and Brr2p unwinds the U4/U6 helix, allowing U6 snRNA to replace the U1 snRNA at the 5' splice site and to base-pair with U2 snRNA (Raghunathan and Guthrie, 1998; Staley and Guthrie, 1999). Prp2p mediates the rearrangement of the first catalytic step of splicing (Chen and Lin, 1990; Kim and Lin, 1996) while Prp16p promotes the second step of splicing (Schwer and Guthrie, 1991; Schwer and Guthrie, 1992). Prp22p has dual roles in pre-mRNA splicing; it is required for the second step of splicing in an ATP-independent manner and release of the mRNA from the spliceosome and initiation of spliceosome disassembly in an ATP-dependent manner (Company et al., 1991; Schwer and Gross, 1998). Finally, Prp43p is involved in the snRNP recycling pathways which are not well understood (Arenas and Abelson, 1997; Martin et al., 2002).

1.2.4 Large nuclear ribonucleoprotein particle, the pre-mRNA processing machine

Consistent with the view of co-transcriptional recruitment of pre-mRNA processing factors, other investigators have shown that mammalian nuclear pre-mRNA is assembled into large nuclear ribonucleoprotein particles (lnRNP), called supraspliceosomes (Azubel et al., 2004; Muller et al., 1998; Sperling et al., 1985) or polyspliceosomes (Wassarman and Steitz, 1993). A remarkable feature of the lnRNP particles is that they package pre-mRNA transcripts of very different sizes and of different numbers of introns; however they are compacted into a complex with a unique size and hydrodynamic properties that sediment at ~200S in sucrose gradients. The lnRNP is thought of as a nuclear pre-mRNA processing machine with all of the U snRNPs required for pre-mRNA splicing, Ser/Arg-rich (Blanton et al.) proteins, A-to-I editing enzymes, cap-binding complex, and 3'-end-processiong components (Raitskin et al., 2002; Yitzhaki et al., 1996). Based on the structure and mass determination by scanning transmission electron microscopy, the lnRNP particle has an average mass of 21.1 MDa and is proposed to have one to eight repeating spliceosomal substructures, each with a mass of 4.8MDa (with a median of four), and the fifth substructure with a mass of 1.9 MDa containing 5'-end capping and 3'-polyadenylation activities (Muller et al., 1998). The pre-mRNA is looped around this huge particle for intron removal and processing. This lnRNP assembly model is hypothesized to result in better efficiency and specificity of recognition of large introns, providing a machine for an accurate and regulated processing of pre-mRNAs.

1.3 AFFINITY TOOLS FOR THE PURIFICATION OF RIBONUCLEOPROTEINS IN YEAST

1.3.1 Application of affinity chromatography for the purification of epitope-tagged proteins

Affinity chromatography is a powerful tool for the purification of biologically active multi-protein complexes, for protein compositional, structural and functional studies. In some cases, such as fluorescent tags, they can also provide information regarding the dynamics and localization of biological molecules inside the cells. In this technique, a fragment of DNA coding for the tag of interest is cloned in frame with the 5' or 3'-terminus of the targeted gene. The expressed fusion protein containing the tag can be affinity-purified by its cognate binding partner immobilized onto a solid-phase matrix. Affinity tags range from small peptides to large proteins. The selection of the appropriate affinity chromatography system and purification condition is based on the nature of the application. One of these affinity tags, the tandem affinity purification tag (TAP tag), has been used to purify multi-protein complexes from yeast with high specificity and efficiency (Puig et al., 2001). The TAP tag contains two affinity cassettes, the *Staphylococcus aureus* protein A (Pro A) and a calmodulin binding peptide (CBP) separated by a TEV (tobacco etch protease) protease cleavage site. Cellular extracts are passed through a non-specific IgG matrix to which Pro A binds tightly for the first step of purification, followed by TEV protease treatment to release the bound material. The eluate is then incubated with calmodulin-coated beads in the presence of calcium. After washing away the contaminants and TEV protease, the purified protein complex is eluted with an EGTA-containing buffer, which chelates the calcium, releasing the complex from the matrix.

1.3.1.1 Gene targeting and affinity chromatography in yeast

In yeast cells, the best and fastest way to generate an epitope-tagged protein is to use a PCR-based genomic tagging approach (Petracek and Longtine, 2002). Briefly, PCR fragments containing the epitope-tag sequence and selectable markers with flanking DNA sequence homologous to the sequence of the targeted locus are transformed into the yeast cells. The highly efficient homologous recombination system in yeast ensures that most of the time, the epitope tag is inserted into the targeted gene in the chromosome at the target site. Yeast cells expressing the epitope-tagged protein are selected on dropout or drug-containing media, depending on the selectable markers. Homologous recombination between transformed DNA and the genome was first observed by two groups, the Beggs laboratory and the Fink Laboratory in the late 1970s (Beggs, 1978; Hinnen et al., 1978). Many factors such as the length of homology at the ends, circular versus linear DNA fragments, and different chromosomal regions influence the frequency of homologous recombination (Rothstein, 1991). Tagged protein expression by gene targeting has been widely used for protein localization and components studies in yeast, on a genomic scale, as well as for the purification of low-abundance RNPs (Ghaemmaghami et al., 2003; Huh et al., 2003; Stevens, 2000).

1.3.1.2 Purification of protein complexes from metazoan cells

In vertebrate and mammalian cells, epitope-tagging in the chromosomal locus is inefficient due to fact that the rate of homologous recombination is extremely low. Analogous purifications of proteins or large complexes are based either on an ectopically expressed trans-gene containing an epitope tag, or the use of the relatively expensive monoclonal antibodies. Generally, the gene of interest is regulated by the human cytomegalovirus (CMV) immediate-early regulatory sequence called CMV promoter.

However, spatial and temporal expression of the native gene locus is lost under an active, constitutive and heterologous promoter; furthermore, the constructs integrate into the chromosome randomly and the site of integration strongly affects the expression level of the epitope-tagged protein. Equally importantly, a single cDNA is used, and the alternative splicing program for the gene of interest is completely lost.

1.3.2 Purification of salt stable yeast snRNPs

Purification and identification of the snRNPs from mammalian cells and yeast were originally based on anti-Sm antibodies, anti-U1 snRNP antibodies, and a monoclonal antibody against the TMG cap portion of the snRNAs. However, due to the low abundance of snRNP proteins in yeast, isolation of intact splicing complexes is extremely difficult. Stevens *et al.* developed a dual affinity chromatography cassette which can be used to purify stable protein complexes from yeast by epitope tagging any gene at its endogenous chromosomal locus (Stevens, 2000; Stevens and Abelson, 1999). The dual affinity cassette consists of a polyhistidine [$(His)_8$] sequence followed by two repeats of the polyoma epitope (Stevens and Abelson, 1999). Yeast U4/U6·U5 tri-snRNP was purified by a 2-step purification using polyoma-agarose chromatography and Ni-NTA resin with subsequent glycerol gradient sedimentation. Yeast U5 snRNP and U6 snRNP were purified by polyoma-agarose chromatography using strain containing a polyoma-tagged *BRR2* gene and *PRP24* gene respectively (Stevens, 2000; Stevens et al., 2001). The newly identified associated factors were further analyzed biochemically and genetically to determine their functional role in splicing (Stevens et al., 2001). Yeast U1 snRNP was isolated by the Lührmann laboratory by anti-TMG-cap immunoaffinity and subsequent nickel-NTA chromatography using a strain containing hexahistidine [$(His)_6$] tagged U1-70K at its N-terminus (Neubauer et al., 1997) and the yeast U2 snRNP

was partially purified in the Seraphin laboratory from TAP tagged U2A' protein (Behrens et al., 1993b; Caspary et al., 1999) Due to the salt-sensitive association of the SF3a and SF3b components (Behrens et al., 1993b) the intact U2 snRNP has never been purified from yeast.

Interestingly, Stevens *et al.* have noticed that during the purification of the yeast U4/U6•U5 tri-snRNP, there are salt-dependent changes in the characteristics of the snRNP complexes (Stevens et al., 2002). When purifying the tri-snRNP from polyoma epitope-tagged Prp4p at 250 mM KCl, only those snRNAs, U4, U5, U6, most tightly associated with the protein were immunopurified. As the salt concentration was lowered to 150 mM, they observed a U2•U4/U6•U5 tetra-snRNP which sedimented at ~ 30S in glycerol gradients. When the salt concentration was reduced to 50 mM during purification, a U1•U2•U4/U6•U5 penta-snRNP was observed which sedimented at ~45S. Based on previous studies, discrete U1, U2, and U4/U6•U5 snRNP particles have been purified under high salt conditions which are not compatible with optimal splicing condition *in vitro* (Lin et al., 1985). Isolated nuclei were extracted under high salt conditions and buffer containing high salt concentration was used to reduce background non-specific binding during the one-step antibody affinity purification. Under stringent conditions, only the most tightly associated protein complexes were immunopurified. It is likely that the proteins that are not present at high salt concentration, but are present under the more splicing-friendly conditions, do participate in splicing.

1.3.3 The discovery of penta-snRNP

Stevens *et al.* showed that the 45S penta-snRNP particle contained 85% of all know yeast splicing factors and exhibited splicing activity when supplemented with micrococcal nuclease (MN) treated extract (Stevens et al., 2002). They proposed that

the penta-snRNP functions as a pre-formed non-dissociating particle by mixing experiments, indicating that endogenous penta-snRNPs are the major precursors to spliceosomes in yeast extracts.

Whether snRNPs join the splicing reaction stepwise or as an assembled complex, or both, is still a debatable issue. However, several groups have evidence suggesting the preassembled particle for splicing. Antibodies against snRNP proteins such as Prp4p (Banroques and Abelson, 1989), Prp6p (Abovich et al., 1990), Prp18p (Horowitz and Abelson, 1993), and Brr2p (Raghunathan and Guthrie, 1998) are able to immunoprecipitate all five snRNAs in stoichiometric amounts under salt conditions around 50-100 mM, optimal for splicing reaction in yeast. However, as the salt concentrations are raised to 200-300 mM, only the RNAs most tightly associated with the proteins are detected. Guialis *et al*. (Guialis et al., 1991) and Moraitou *et al*. (Moraitou et al., 1998), separately showed a penta-snRNP-like particle from mammalian cells that contains all five snRNAs.

The Sperling group isolated a native spliceosome complex from supraspliceosomes by hybridizing the supraspliceosome with a DNA oligonucleotide complementary to the consensus 5' splice site sequence, followed by RNase H digestion (Azubel et al., 2006; Azubel et al., 2004). Visualization by cryo-electron microscopy, and 3D model reconstruction showed a globular shape consisting of 2 distinct subunits, one large and one small, connected to each other leaving a tunnel in between (Azubel et al., 2004). Based on the mass and volume calculation, the larger subunit is a suitable candidate for the penta-snRNP particle while the small subunit is proposed to contain other pre-mRNA processing factors and the tunnel is where the pre-mRNA passes through. They further proved that the native spliceosome, containing all five snRNAs, is splicing competent (Azubel et al., 2006).

1.4 DISSERTATION OBJECTIVES

Epitope tagging technology has enhanced the tool kit of the molecular biologist by allowing the facile detection, tracking and purification of virtually any polypeptide in any model organism. Adding epitope tags to proteins is generally accomplished in metazoan cells using ectopically expressed, trans-genes. I began my dissertation work with the intention of developing a better biochemical tool for purification which circumvents the shortcomings of trans-genes by introducing the epitope tag into the native chromosomal gene locus, termed CLEP tagging (Chromosomal Locus EPitope tagging) in vertebrate cells, embryonic stem cells and live mice. This method allows us to purify tagged proteins coded by genes controlled by their native promoter and with the use of different tags, facilitating the localization and function of the protein of interest.

This dissertation comprises five chapters. Chapter 1 provides an overview of our current understanding of the gene expression network with a particular focus on pre-mRNA splicing machinery and the application of epitope-tagging method on characterization of snRNPs. Chapter 2 contains the materials and methods of CLEP tagging technique by inserting a TAP tag or histidine (His) tag into the 3' terminus of the SmD3, SF3b155 and Lsm3 genes in cultured vertebrate cells. I show that this method of purification is robust and sensitive by identification of low-abundance proteins and novel proteins involved in splicing that have not been characterized in metazoan cells before. In Chapter 3, I describe that purifying *in vivo*-assembled pre-mRNA processing machines from DT40 cells and HeLa cells expands the catalog of participating factors. With a similar CLEP technique, a TAP tag and a polyhistidine tag insertion into the 3' terminus of SART1 gene and U5-220K gene, respectively, in mouse embryonic stem cells and generation of mice processing a TAP tagged SART1 gene, are described in Chapter 4. In the last Chapter, I demonstrate that modifying SART1 protein with a TAP

tag triggers an autoimmune response in mice which contain a TAP tag at both loci. These mice will serve as a valuable tool for the studies of mammalian autoimmune diseases. In summary, the improvement of epitope tagging and purification technique extends our understanding of the complex gene expression network.

Chapter 2: Gene targeting in cultured vertebrate cells

2.1 BACKGROUND

In order to create a biochemical tool in a vertebrate system which circumvents the shortcomings of the use of trans-genes, I extended the use of homologous recombination technology in appropriate cell lines capable of efficient homologous recombination techniques.

2.1.1 A vertebrate cell line useful for gene targeting

The DT40 chicken pre-B cell line is derived from avian leucosis virus (ALV)-induced bursal lymphoma, expressing IgM on its surface (sIgM). These cells have integrated the 3' viral long terminal repeat upstream of the c-myc gene, resulting in the transcription of c-myc at levels 100-fold higher than normal avian bursal cells (Baba et al., 1985). Importantly, the c-myc gene is deregulated by adjacent retroviral integration in these lymphocytes and maintains them in a stem-cell stage (Thompson et al., 1987). Deregulation of c-myc gene during B cell development in the Bursa blocks differentiation and leads to the expansion of the bursal stem cell population. Baba *et al*. and Kim *et al.* further reported that these cell lines possessed distinct c-myc alleles; some contained one normal and one deregulated c-myc while others lacked normal c-myc genes, suggesting these cells generated a mitotic recombination event among the parental alleles (Baba et al., 1985; Kim et al., 1990). DT40 is also an ideal model for the study of somatic diversification mechanisms after immunoglobulin gene rearrangement as DT40 cells continue to undergo gene conversion as shown by testing for the presence of absence of restriction sites within the light chain variable region (Thompson et al., 1987).

In 1991, Buerstedde *et al*. reported that DT40 cells exhibited high frequencies of targeted integration several orders of magnitude higher than the frequencies reported for transfection of mammalian cells (Buerstedde and Takeda, 1991). After transfection with a construct containing a neomycin resistance gene (neo^R) into the immunoglobulin gene locus, the DT40 cells lost their surface IgM expression due to the targeted integration. The high frequency of targeted integration was neither restricted to the gene conversion activity within the light chain locus nor restricted to the untranscribed genes, as they tested targeted integration into the β-actin locus (unrelated to light chain locus) and ovalbumin locus (not transcribed in chicken B cells). This cell line offers a targeted DNA integration frequency of 10% to 90%, depending on the locus of interest.

The use of the DT40 cells is increasing in popularity because of the ease with which it can be manipulated in suspension tissue culture media, the high rate of homologous recombination, a rapid (8-hour) doubling time, and exceptional conservation with mammalian cells of cellular processes. The high frequency of targeted integration is a general feature of DT40 and it has been widely used in gene disruption experiments, contributing to our understanding of diverse processes including B cell antigen receptor signaling, apoptosis, histone gene function, RNA processing, DNA repair, and cell cycle [reviewed in (Winding and Berchtold, 2001)].

2.1.2 A human cell line useful for gene targeting

Only a few reports on successful gene targeting in human cell lines are available. Unfortunately, none of them were shown to be more recombinogenic than the others. In 2004, Feederle *et al.* reported a human EBV-negative Burkitt lymphoma B-cell line, DG75, capable of efficient homologous recombination similar to that of the DT40 system (Feederle et al., 2004). They generated a homozygous knockout of a cellular gene

encoding a protein interacting with a viral origin of replication and showed that genetic manipulation of DG75 cells was relatively easy. This report provided a useful tool for studying function through gene targeting in human cells.

In this chapter, I take an advantage of the efficient homologous recombination in the DT40 cell line to insert an epitope-tag in the chromosome locus of a gene by homologous targeting. I employ chicken DT40 pre-B cells to TAP- or $(His)_8$-tag the SmD3, SF3b155 or Lsm3 genes in this model vertebrate cell line (Buerstedde and Takeda, 1991). Tagging these ribonucleoprotein components allows us to perform proteomic and functional analyses of the splicing machinery in metazoans.

2.2 MATERIALS AND METHODS

2.2.1 The design of targeting vector

The general strategy for the CLEP tagging procedure is shown in Figure 2.1. To create the targeting vector, three DNA regions from the gene of interest were amplified by polymerase chain reaction (PCR) from genomic DNA purified from DT40 cells. The AB fragment (Figure 2.1B) is the region of 1 – 4 Kb upstream of the stop codon of the gene of interest containing a BamHI or BglII site immediately upstream of the STOP codon. The 5' end of the AB fragment is generally designed to be contained in an intronic sequence. The CD fragment (Figure 2.1B) is typically the 1 Kb of genomic DNA downstream of the STOP codon of the gene of interest, although this may be made longer if the sites of cleavage polyadenylation are known to be further downstream that 1 Kb from the STOP codon. The EF fragment is the region 2-4 Kb further downstream of the CD fragment. All fragments were PCR amplified (Expand High Fidelity PCR kit, Roche Applied Science) from the genomic DNA purified from the DT40 cells and cloned in the pGEM-T-Easy (Promega) vector. All fragments were sequenced to verify that the

coding region and splice site signals in AB have been maintained and to verify the identity of the CD and EF fragments. The AB, CD, and EF fragments for each gene were amplified using oligonucleotides as shown in Table 2.1.

The targeting vector backbone we used is pOSDUPdel (a kind gift of W. Kuziel), a positive-negative selection vector, containing a neomycin phosphotransferase gene (*neo*) cassette flanked by loxP sites and a thymidine kinase gene conferring ganciclovir resistance. The restriction enzyme sites used in these constructions are flexible. The AB and CD fragments are cloned into the first polylinker region containing the following sites, in order: NotI, PmeI, XbaI, HinDIII, KpnI, HpaI, SalI, BamHI and PacI. The EF fragment is cloned into the second polylinker region containing the following sites in order: ClaI, XhoI, PmlI, BclI, and NheI. For simplicity, NotI and SalI are generally used for cloning of the AB fragment and SalI and PacI for the cloning of the CD fragment as it eliminates the BamHI site in the polyliner such that a BamHI fragment containing the epitope tag can be inserted or replaced as needed. Alternatively, the epitope tag-containing BamHI fragment can be added to the AB fragment in the pGEM-T-Easy vector. The cloning strategy should take into account the order in which the fragments are inserted into the targeting vector, and ensure that the sites used for cloning are absent from the fragments yet to be inserted. The epitope tag, TAP or $(His)_{8,}$ is contained on a BamHI restriction fragment and cloned into the BamHI or BglII site incorporated into the AB fragment.

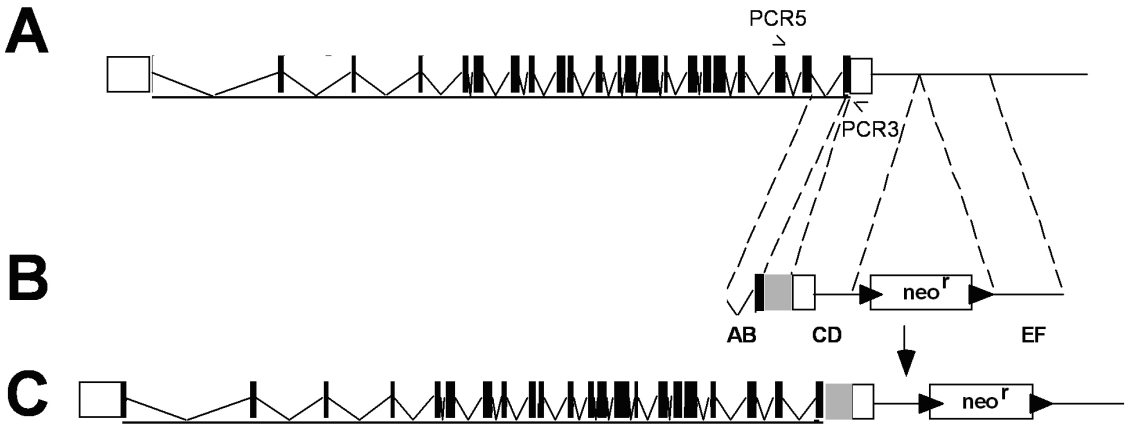

Figure 2.1 Strategy for the epitope tagging technology in DT40 cells, ES cells and live mice.

(A) Idealized gene structure of a metazoan gene. Exons are represented by solid rectangles, introns by thin lines. The epitope tag is represented by a grey rectangle. (B) Targeting vector fragment containing the sequences required for the introduction of the epitope tag, and the regions of homology for the targeting of the gene. (C) Idealized gene sequence with targeting construct inserted. Locations of oligonucleotides used in PCR and RT-PCR (PCR5 and PCR3) are noted by arrows.

Table 2.1 Oligonucleotides used for targeting vector construction in DT40 cells

Targeted gene	Oligo sequence
GgLSM3-A	5'-GCGGCCGCGGTGAGGACAGATCCACGATGCACTGG-3'
GgLSM3B-HIS	5'-GTCGACGGATCCTTAGTGGTGATGATGATGATGATGATGGG ATCCGCC AACCCTCAGTGGGGGAGCTACAAGCAC-3'
GgLSM3-C	5'-GTCGACAGCAACCAAGGATTGAACCTTCTTGGAAG-3'
GgLSM3-D	5'-TTAATTAAGGATCCTAGGCAAGGTATGTCAAGAAATGCCTG-3'
GgLSM3-E	5'-CTCGAGGGATCCGGTTTTAGACACGTTCTGTACAATCATC-3'
GgLSM3-F	5'-GCTAGCGCAACTTCTGGGCAAAAAGGGAAGGCAAATG-3'
GgSF3b155-A	5'-GCGGCCGCCTTCCAACCCAAACATTCTGTGATTCTC-3'
GgSF3b155-B-HIS	5'-GTCGACGGATCCTTAGTGGTGATGATGATGATGATGATGGG ATCCTAAGATGTAGTCAAGTTCATAACGAATATAGG-3'
GgSF3b155-C	5'-GTCGACTCTTCTTGTTCTGTTGTTTGTGTTTAATGC-3'
GgSF3b155-D	5'-TTAATTAAGGATCCAACACTACCAGAATGACGTAACCTTGTG-3'
GgSF3b155-E	5'-CTCGAGGGATCCGGGTAGTGTTTCATGTCCATAAGAAAAC-3'
GgSF3b155-F	5'-GCTAGCGTACTTTCCAAGACAGTTTAGAGTTTGGCAG-3'
Gg SmD3-A	5'-GCGGCCGCAGGCTGTATGCTTGACAGGGCTTTGAG-3'
GgSmD3-B	5'-GTCGACAGATCTTCTTCGCTTCTGGAAGATGTTGCCACGACC-3'
Gg SmD3-C	5'-GTCGACGACCAGTATGCTTTTTTTTATTAGAGG-3'
Gg SmD3-D	5'-TTAATTAAGGATCCTCATTCATACATGTATAGACTGATGAC-3'
Gg SmD3-E	5'-CTCGAGGGATCCTATAAAGACCCCCTGTGTGCAGACATG-3'
Gg SmD3-F	5'-GCTAGCCCTTCTATCCACAGGTGTGATCTTCC-3'

2.2.2 DT40 cell culture and targeting vector transfection

2.2.2.1 Growth condition and media preparation

DT40 cells were grown in Dulbecco's modified Eagle media (DMEM) supplemented with 5% chicken serum, 5% fetalplex (Gemini Bio-Products), and antibiotics (penicillin/streptomycin, 100X, Gibco/invitrogen). When cultured at 37℃ with 8% CO_2, the generation time for the wild-type cells is about 8 hours.

2.2.2.2 Cryopreservation of cells

The cells were grown in 30 mL medium to confluency (~10^6 cells/mL), followed by centrifugation for 5 minutes at 1000 rpm. The cell pellets were thoroughly resuspended in cryopreservation media: 2 mL solution A (16 ml of DMEM with 4 mL of fetalplex) and 2 mL solution B (15 ml of DMEM, 6 ml fetalplex and 9 ml of DMSO); aliquot 1 mL into each of 4 labeled cryovials. The cryovials were stored in styroform boxes in -80˚C for 2 days and then moved into liquid nitrogen.

2.2.2.3 Recovery of frozen cultured cells

The cryovials were thawed in a 37˚C water bath, with frequent shaking. Thawed cells were transferred into 5 mL DMEM media, and centrifuged for 5 minutes at 1000 rpm. The pelleted cells were resuspended in 10 ml of DMEM media. Cells were grown in a T25 flask to confluence.

2.2.3 DNA transfection into DT40 cells

After the targeting vectors containing the epitope tagging fragment were created, the plasmid DNA (20-30 µg) was linearized at a unique site (generally NotI or NheI) and electroporated into the DT40 cells. Generally, 10^7 DT40 cells, in 300 µl of phosphate buffered saline, PBS , (137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4) were transferred into cuvettes and electroporated with 20 µg linearized targeting constructs (BioRad Genepulser II, 700 V, 25µF, 0.4 cm cuvette). After electroporation, the cells were recovered in 15 ml non-selective media for 24 hours, then harvested and resuspended into 40 ml of selective media containing 1.5 mg/ml of G418 for neomycin selection. 100 µl aliquots were distributed into each well of four 96 well plates. G418-resistant cells formed colonies after 10-14 days in selective media. They were expanded into 48 well plates and grown in G418 selection media until confluent. RT-PCR or

western blot analyses were performed to confirm the presence of the epitope-tag insertion into the correct chromosomal locus.

2.2.4 Screening for tagged gene by RT-PCR analysis

For screening DT40 cells positive for SmD3-TAP, Lsm3-$(His)_8$, SF3b155-$(His)_8$, total RNA was isolated (RNAwiz, Ambion) from 10^5 cells for RT-PCR analysis using oligonucleotides as shown in Table 2.2. RNA was resuspended in 30 µl RNase free-water and 50ng of RNA was used for one step RT-PCR (One Step RT-PCR kit, Qiagen).

Table 2.2 Oligonucleotides for RT-PCR analysis in DT40 cells

Targeted gene	Oligo sequence
Gg LSM3PCR5	5'-ACGCATATGATCAGCATTTAAATATGATTC-3'
Gg LSM3PCR3	5'-AGCCTGAAACCTTCCAAGAAGGTTCAATCC-3'
GgSF3b155PCR5	5'-TTGCAGTATTGTTTGCAGGGTTTGTTTCAC-3'
GgSF3b155PCR3	5'-TTA TGTGAAGAACAGCTGTGCATTAAACAC-3'
GgSmD3PCR5	5'-CGAGGAAAAGCAGCTATTCTCAAAGCTCAG-3'
GgSmD3PCR3	5'GACAGAAGTTACCAACATATATGTAGAC-3'

2.2.5 Screening for tagged gene by western blot analysis

10^5 cells were lysed by addition of 200 µl LDS-PAGE loading buffer [4X sample buffer (pH 8.5): 106 mM Tris-HCl, 141 mM Tris base, 2% LDS 10% glycerol, 0.51 mM EDTA, 0.22 mM SERVA Blue G250, 0.175 mM phenol red] and homogenized samples (20 µl) were electrophoresed through 10% SDS-PAGE gels and transferred to nitrocellulose membrane. Membranes were blocked for 1 hour in PBST (PBS + 0.2% Tween 20) containing 5% dry milk solution. Membranes were then incubated with Peroxidase-antiperoxidase (PAP) antibody (Sigma) (1:1000 dilution) for 1 hour to detect

TAP tagged proteins. After washing membranes extensively with PBST, the signal was detected by enhanced chemiluminescence (Perkin Elmer).

2.2.6 Extract preparation

Six liters of SmD3-TAP DT40 cells were grown in Dulbecco's modified Eagle media supplemented with 5% chicken serum and 2.5% Fetalplex to a density of 7.5×10^5 cells/ml for TAP purification. Cells were harvested by centrifugation (400 x g for 5 minutes), washed twice with ice cold PBS, allowed to swell in 10 ml of TM buffer [10 mM Tris-HCl, (pH 7.5), 3 mM $MgCl_2$] with 0.2 mM PMSF, 1 µg/ml leupeptin, and 1 µg/ml pepstatin for 10 minutes on ice, and lysed with 25 strokes of a Dounce homogenizer at 4˚C. The nuclei were pelleted, washed twice with 10 ml of TM buffer containing 0.1% NP40, resuspended in 5 ml of low salt buffer (30 mM Tris-HCl, 125 mM KCl, 5 mM $MgCl_2$, 0.5% Triton-X100, 0.15mM spermine, 0.05mM spermidine, and 0.2 mM PMSF), and sonicated at the maximum output, twice for 20 seconds on ice with 1 minute in ice between sonications. The sonicated mixture was centrifuged at 5000 rpm for 10 minutes and the supernatant, the nuclear fraction, was used for TAP purification.

2.2.7 TAP purification procedure

TAP-tagged nuclear protein material for SmD3-TAP DT40 cells was affinity purified by the TAP procedure (Puig et al., 2001). The nuclear fraction was incubated with 400 µl of IgG sepharose beads for 2 hours at 4˚C in buffer IPP150 (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% NP40). The tagged protein and associated material were recovered from extracts by IgG affinity selection. After washing with 100 column volumes of IPP150, the beads were suspended in TEV cleavage buffer (IPP150 containing 1 mM DTT). TEV protease (2 µl for 400 µl bead volume) was added to

release the bound material at 16˚C for 2 hours. The eluate was then incubated with 200 µl calmodulin-coated beads in IPP150 calmodulin binding buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% NP40, 10 mM 2-mercaptoethanol, 1 mM magnesium acetate, 1 mM imidazole, 2 mM $CaCl_2$) and rotated for 1 hour at 4˚C. The calmodulin-coated beads were washed with 100 column volumes of IPP150 calmodulin binding buffer, and the bound material was released with IPP150 calmodulin elution buffer (10 mM Tris-HCl pH 8.0, 150 mM NaCl, 0.1% NP40, 10 mM 2-mercaptoethanol, 1 mM magnesium acetate, 1 mM imidazole, 4 mM EGTA). Five elution fractions of 200 µl each (1 bead volume) were collected. RNA and protein were separated by phenol/chloroform extraction. RNA samples from aqueous phase were precipitated by addition of 0.3 M NaOAc, pH 5.3 and followed by addition of 2.5 sample volumes of 100 % ice-cold ethanol, assayed on 7% polyacrylamide gel (19:1), and visualized by ethidium bromide or silver stain. Protein samples were precipitated by 5 sample volumes of ice-cold acetone from the organic phase, assayed on 10% SDS-PAGE gel, and visualized by staining with silver or Coomassie blue G-250.

2.2.8 Mass spectrometry analysis

SDS-PAGE separated affinity purified polypeptides were excised from the gel analyzed by mass spectrometry (LS/MS/MS) as previously described (Davis and Lee, 1998; Moore et al., 2000). Fragmented ion mass spectra were screened and searched against the NCBI non-redundant database using the Sequest matching program. The Qscore scoring procedure was used to measure the quality of the results for protein identification obtained from the peptide fragment matching of the MS/MS data using the Sequest database program (Moore et al., 2002). Scoring system is based on the estimation of the probability that protein identification was by chance. Based on the

number of identified peptides from the protein, the total number of identified peptides, and the percentage of distinct tryptic peptides from the database covered in the identified proteins, the Qscore is assigned to each identified proteins. This scoring system allows us to eliminate contaminants from the Sequest protein identifications. Consequently, each individual spectral fragment is validated manually to ensure that each peak matches the molecular weight of proteins on the gels.

2.3 RESULTS

2.3.1 DT40 cells carrying CLEP tags

I employed the chromosomal locus epitope tagging procedure in a model vertebrate system, the DT40 chicken pre-B cell line. I chose three target genes in DT40: the SmD3, SF3b155 and Lsm3 genes encoding polypeptides participating in pre-mRNA splicing and/or RNA turnover. The design of the targeting vectors was performed as described above. After construction and clone validation, the targeting vector was linearized at an appropriate site upstream of the AB fragment (usually NotI) or downstream of the EF fragment and electroporated into DT40 cells. The cell colonies that survived in the presence of G418 selection media were chosen and expanded. The presence of correctly targeted cells was tested by PCR using one oligonucleotide primer designed outside the targeting region (PCR5 in Figure 2.1) and one oligonucleotide primer designed inside the targeting vector (PCR3 in Figure 2.1). Lsm3-$(His)_8$ was positively identified in 19/30 colonies (Figure 2.2), SF3b155-$(His)_8$ was identified in 12/30 colonies (Figure 2.2) and SmD3-TAP was recovered in 2/20 colonies (Figure 2.3A and 2.3B). I note that the smaller size of the $(His)_8$ tag is a likely reason they were more efficiently recovered than TAP-tagged constructs (10% of positives in the TAP-tagged construct vs. 63% and 40% for HIS-tagged constructs). However, I did not try to insert

different tags into same chromosomal locus. One-step purification using Ni-NTA affinity chromatography to purify SF3b155-$(His)_8$ and Lsm3-$(His)_8$ associated components was not efficient or pure from chicken cells so I did not proceed with further experiments with these two tagged cell lines.

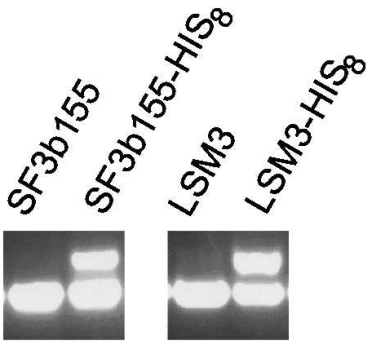

Figure 2.2 Targeted insertion of $(His)_8$ tags into DT40 genes.

RT-PCR analysis of SF3b155-$(His)_8$ and Lsm3-$(His)_8$ targeting construct insertion into the respective DT40 genes. Lower band represents the wild-type allele and upper band represents the targeted allele.

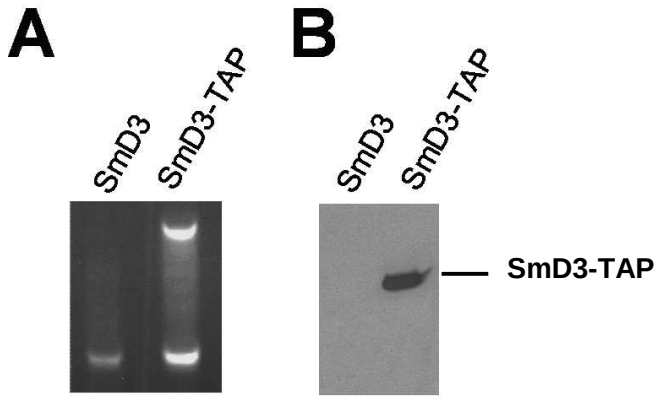

Figure 2.3 Targeted and functional insertion of TAP tag into DT40 genes.

(A) RT-PCR analysis of SmD3-TAP targeted clones. Lower band represents the wild-type allele and upper band represents the targeted allele. (B) Western blot analysis of SmD3-TAP clones. SmD3-TAP polypeptide is noted.

2.3.2 Identification of SmD3-associated polypeptides

To demonstrate functional association of the tagged material with its native cellular machinery, 6 liters of SmD3-TAP DT40 cells were grown to a density of $7.5x10^5$ cells/ml. Cells were harvested and nuclei were fractionated and disrupted by sonication. SmD3-TAP-associated material was purified by the TAP procedure (Puig et al., 2001) and associated polypeptides were separated by SDS-PAGE (Figure 2.4) and analyzed by mass spectrometry peptide analysis (Davis and Lee, 1998). SmD3-associated polypeptides from selected bands are shown in Table 2.3. The identified polypeptides correlate to most of the known SmD3-associated spliceosomal snRNP proteins (Jurica and Moore, 2003) demonstrating that the SmD3-TAP polypeptide is

being faithfully incorporated into functional complexes and includes factors shown to be only loosely associated (Will et al., 2002)(e.g. SR140, SAP45 and SAP30). Indeed even very low-abundance U11/U12 snRNPs (U11/U12-65K and U11/U12-48K) were identified in the salt-stable snRNP fractions demonstrating the sensitivity of the procedure.

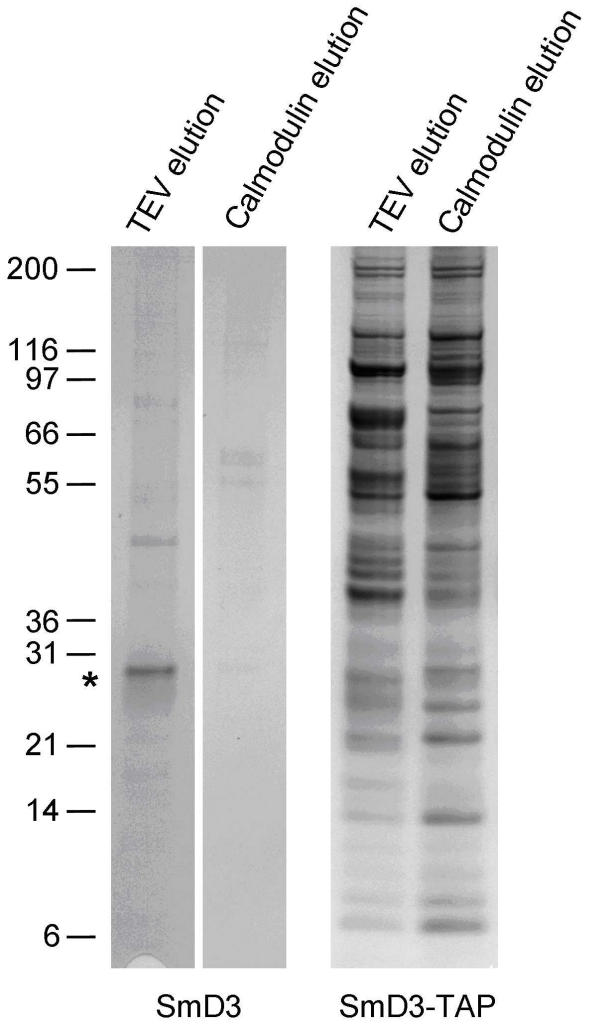

Figure 2.4 Functional incorporation of SmD3-TAP into splicing snRNPs.

Untagged DT40 and tagged SmD3-TAP cell nuclei were subjected to TAP purification procedures (Puig et al., 2001) and affinity-purified polypeptides resolved by SDS-PAGE. TEV elution and calmodulin elution profiles are shown for each. An asterisk marks the location of the TEV protease.

Table 2.3 SmD3-associated polypeptides

Identified protein	ENSEMBL name	snRNP association
U1-70K	ENSG00000104852*	U1
U1C	ENSGALG00000002773	U1
U1A	ENSGALG00000008729	U1
SF3b155	ENSGALG00000008038	U2
SF3b145	ENSG00000087365*	U2
SF3b130	ENSGALG00000002531	U2
SF3b125	ENSGALG00000000581	U2
SF3a120	ENSGALG00000007937	U2
SF3a60	ENSGALG00000001540	U2
DDX15/Prp43	ENSGALG00000014395	U2
SR140	ENSGALG00000002612	U2
SPF45	ENSGALG00000006332	U2
SPF30	ENSGALG00000008561	U2
U2B''	ENSGALG00000008729	U2
U2A'	ENSGALG00000007170	U2
SF3b14	ENSGALG00000016501	U2
U5-220K	ENSGALG00000002943	U5, U4/U6•U5
U5-200K	ENSGALG00000015465	U5, U4/U6•U5
U5-116K	ENSGALG00000000988	U5, U4/U6•U5
U5-110K	ENSG00000175467*	U5, U4/U6•U5
U5-102K	ENSGALG00000006001	U5, U4/U6•U5
U5-100K	ENSG00000174243*	U5, U4/U6•U5
U5-40K	ENSGALG00000000615	U5, U4/U6•U5
U5-15.5K	ENSGALG00000017396	U5, U4/U6•U5
U4/U6-90K	ENSGALG00000000465	U4/U6, U4/U6•U5
U4/U6-61K	ENSG00000105618*	U4/U6, U4/U6•U5
U4/U6-60K	ENSGALG00000008857	U4/U6, U4/U6•U5
U4/U6-20K	ENSGALG00000004874	U4/U6, U4/U6•U5
U4/U6-15K	ENSGALG00000011931	U4/U6, U4/U6•U5
U4/U6•U5-65K	ENSG00000168883*	U4/U6•U5
SmB/SmB'	ENSGALG00000007250	U1, U2, U4/U6, U5
SmD3	ENSGALG00000006596	U1, U2, U4/U6, U5
SmD2	ENSG00000125743*	U1, U2, U4/U6, U5
SmD1	ENSGALG00000011842	U1, U2, U4/U6, U5
SmE	ENSGALG00000000137	U1, U2, U4/U6, U5
SmF	ENSGALG00000011409	U1, U2, U4/U6, U5
SmG	ENSG00000143977*	U1, U2, U4/U6, U5
SART3	ENSGALG00000004887	U6, U4/U6
Lsm2	ENSG00000111987*	U6, U4/U6, U4/U6•U5
Lsm6	ENSGALG00000009985	U6, U4/U6, U4/U6•U5
Lsm8	ENSGALG00000009110	U6, U4/U6, U4/U6•U5
U11/U12-65K	ENSGALG00000005162	U11/U12
U11/U12-48K	ENSGALG00000013005	U11/U12
Prp38	ENSGALG00000010627	Yeast U4/U6•U5**

Table 2.3 (*Continued)*

Identified protein	ENSEMBL name	snRNP association
Prp39	ENSGALG00000012468	Yeast U1**

* Not annotated for *Gallus gallus*. Identification was based on database calls to orthologues from other species.

** Not previously shown to be snRNP-associated in metazoans

2.3.3 Previously uncharacterized metazoan snRNP polypeptides associated with SmD3 in chicken cells

The yeast genes *PRP38* and *PRP39* encode polypeptides that are associated with the yeast U4/U6•U5 and U1 snRNPs, respectively (Gottschalk et al., 1998; Stevens and Abelson, 1999). These two polypeptides have never been shown to be stable components of metazoan snRNPs when purified using a monoclonal antibody. However, they were identified in our purification, confirming the robustness of the CLEP tagging purification procedure. Sequence alignments of the vertebrate homologues of the yeast Prp38p and Prp39p are presented in Figures 2.5 and 2.6, respectively.

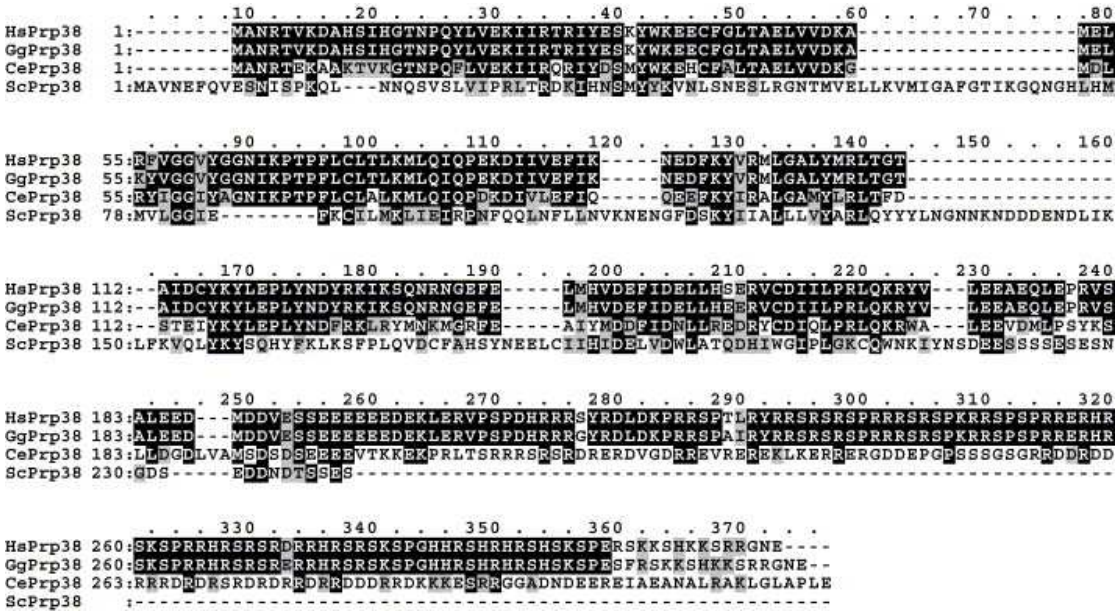

Figure 2.5 Sequence alignments of polypeptide homologues of yeast Prp38p.

Sequences from Human (HsPrp38), chicken (GgPrp38), and worm (CePrp38) were aligned with yeast Prp38p polypeptides. Identical amino acid residues are highlighted with white text on a black block. Similar amino acid residues are highlighted with black text on a gray block.

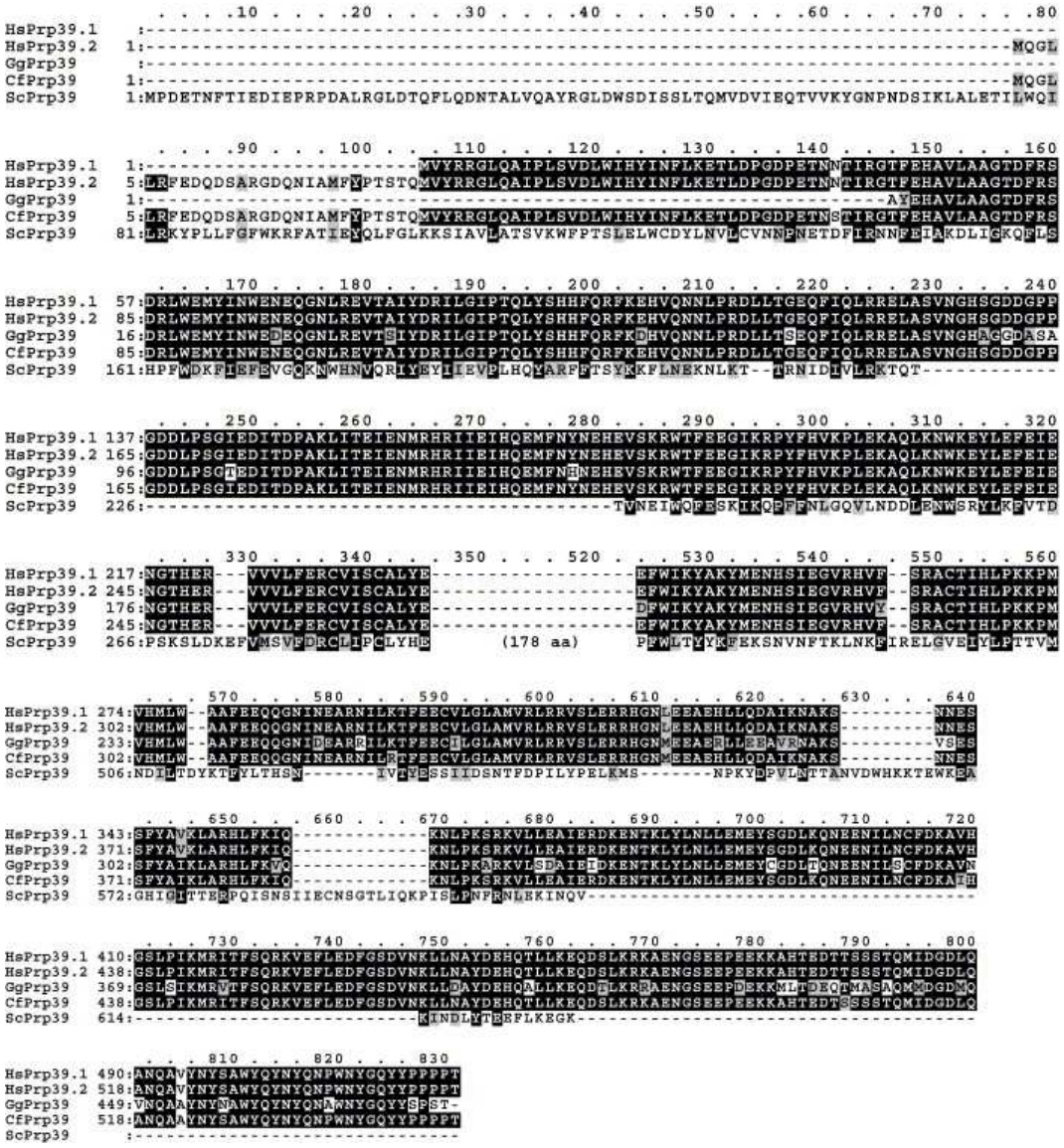

Figure 2.6 Sequence alignments of polypeptide homologues of yeast Prp39p.

Sequences from Human (HsPrp38), chicken (GgPrp39), and worm (CePrp39) were aligned with yeast Prp38p polypeptides. Identical amino acid residues are highlighted with white text on a black block. Similar amino acid residues are highlighted with black text on a gray block.

2.4 DISCUSSION

I have demonstrated the feasibility and promise of incorporating epitope tags into the chromosomal loci of essential genes in cultured vertebrate cells. I also have shown that the procedure is robust and highly sensitive, as loosely associated factors and extremely low-abundance material such as the U11/U12 snRNP proteins, were purified and identified by mass spectrometry. Indeed a comparison between conventional antibody chromatography and the use of the CLEP tag shows a much greater number of polypeptides identified by mass spectrometry using the CLEP tag.

With tissue culture cells, once a tagged cell line has been established using CLEP, an unlimited supply of cells can be grown for experimentation for which precious and expensive antibodies were previously required. Purification of protein complexes controlled under its native promoter enhances the validity of conclusions made regarding epitope-tagged proteins versus the use of a trans-gene. Furthermore, the use of different epitope tags on multiple genes will not only allow additional flexibility and experimentation, such as co-immunoprecipitation analysis, but also offer a useful method for the separation of subsets of complexes in which a polypeptide of interest exists in multiple complexes with mutually exclusive binding partners. Along with recent and future progress in genome sequencing, these experiments will be easily applicable to other metazoan organisms for which extensive genomic sequencing has been performed, and for which cells possessing efficient homologous recombination are available.

Chapter 3: Proteomic analysis of the endogenous pre-mRNA processing machinery

3.1 BACKGROUND

Pre-mRNA splicing is a fundamental step in eukaryotic gene expression before the mature mRNA is exported from the nucleus to the cytoplasm where it is translated into protein. The machinery that splices pre-mRNA is called the spliceosome which is comprised of hundreds of proteins and five small nuclear RNAs (U1, U2, U4, U5, and U6) assembled into small nuclear ribonucleoproteins (snRNPs). In addition to the snRNP specific proteins, a large number of spliceosome-associated proteins were identified through biochemical or genetic studies. These proteins include DEAD-box helicase family members, RNA binding proteins, SR proteins, and spliceosome associated proteins, which are important for regulating the many temporally and spatially important RNA and protein rearrangements during the splicing reaction (Staley and Guthrie, 1998).

In an alternative systematic approach to the traditional characterization of the individual snRNP, several groups have purified and characterized spliceosomes using a large-scale proteomic analysis (Neubauer et al., 1998; Rappsilber et al., 2002; Zhou et al., 2002). These purifications involved the assembly of spliceosomes from salt-extracted HeLa nuclei onto synthetic single-intron pre-mRNAs *in vitro* to purify spliceosomes from a mixture of all stages of assembly. Zhou *et al*. assembled spliceosome on AdML-M3 (adenovirus major late) pre-mRNA containing three hairpins that bound to the MS2-MBP fusion protein used for affinity chromatography. Neubauer *et al*. and Rappsilber *et al*. used biotin-labeled AdML and β-globin transcripts as substrates for the spliceosome assembly followed by streptavidin bead affinity selection. Combined with the

improvements in mass spectrometric methods and expanding gene databases, these three comprehensive proteomic analyses of the human spliceosome revealed a remarkably large number of polypeptides associated with splicing complexes.

Nevertheless, native pre-mRNAs contain multiple, often extremely large introns; the *in vitro* pre-mRNA splicing reactions were carried out using synthetic pre-mRNA fragments containing a single, efficiently spliced intron of a size compatible with acrylamide gel electrophoresis analysis. Although the core pre-mRNA processing machinery will likely be very similar between different transcripts as well as for the multiple introns contained within a single transcript, the bulk pre-mRNA processing machinery purified from its native context is likely to contain a more comprehensive sample of the polypeptides required for or participating in the splicing of pre-mRNA in vertebrate cells. In addition, the spliceosomes purified from *in vitro* reactions were assembled on pre-mRNAs derived from either the adenovirus major late or β-globin loci. Thus, it is likely that there exist a number of factors that are required for or participate in pre-mRNA processing *in vivo*, yet are not present in previously purified splicing complexes because they are specific to one or more of the thousands of other pre-mRNAs present in metazoan cells. Finally, the pathway by which pre-mRNA processing complexes are assembled *in vitro* using salt-extracted nuclear fractions most likely bypasses many interactions relevant to this process *in vivo.* Thus, spliceosomes purified following *in vivo* assembly are expected to contain additional components that reflect the native pathway, but are not required to effect model intron removal *in vitro*.

Consistent with this view, other investigators have shown that endogenous pre-mRNA is processed in extremely large ribonucleoprotein particles, called supraspliceosomes (Azubel et al., 2004; Muller et al., 1998; Sperling et al., 1985) or polyspliceosomes (Wassarman and Steitz, 1993), which sediment at ~200S in sucrose

gradients. Biochemical and structural analyses of these complexes have demonstrated the presence of RNA Pol II transcripts and the pre-mRNA splicing machinery as well as functional interactions that mirror those in active splicing complexes assembled *in vitro* (Wassarman and Steitz, 1993). The higher-order particles formed *in vivo* partly reflect the presence of multiple introns, an average of eight per pre-mRNA (Lander et al., 2001) with some transcripts possessing as many as 147 introns [Nebulin (Sakharkar et al., 2004)], that need to be faithfully removed prior to nuclear export.

With the goal of expanding our understanding of pre-mRNA splicing as it occurs in intact cells, I have purified the endogenous pre-mRNA processing machines from chicken DT40 pre-B cells on a preparative scale and have defined their RNA and polypeptide compositions in this chapter. The results from chicken DT40 cells are compared with that from HeLa polyspliceosome purified by conventional means, using Y12 antibodies against Sm proteins (a gift of Joan Steitz). I chose the chicken DT40 system to compare to the HeLa system for a number of reasons. First, I have shown that this rapidly growing cell type, which possesses high rate of homologous recombination, allows for efficient epitope-tagging of genes. The evolutionary distance between human and chicken will allow me to assess the evolutionary conservation of the machinery as well as validating novel co-purifying factors. I show that these pre-mRNA processing complexes contain spliced and unspliced mRNAs, all five spliceosomal snRNAs and polypeptides involved in all aspects of pre-mRNA processing from transcription to nuclear export.

3.2 MATERIALS AND METHODS

3.2.1 Purification of human supraspliceosome from HeLa cells

Ten liters of HeLa cells (purchased from the National Cell Culture Center) were processed essentially as previously described (Sperling et al., 1985). Briefly, cells were washed in PBS and disrupted by mechanical breakage in a glass dounce (20 strokes, pestle "B") in a hypotonic solution (30 mM Tris-HCl pH 7.5, 10 mM KCl, 5 mM $MgCl_2$, 10 mM 2-mercaptoethanol) at 4˚C. Nuclei were pelleted at 1000 x g at 4˚C for 5 minutes through the hypotonic buffer containing 25% glycerol. The nuclei were washed three times in the hypotonic buffer containing 0.5% Triton X-100 and once with detergent-free hypotonic buffer. Nuclei were resuspended in a low-salt buffer (LS+; 10 mM Tris-HCl pH 7.5, 100 mM KCl, 2 mM $MgCl_2$, 10 mM 2-mercaptoethanol, 0.15 mM spermine, 0.05 mM spermidine) and sonicated twice for 20 seconds at the maximum microtip setting. The resulting nuclear debris was pelleted at 14,000 RPM for 10 seconds, and the supernatant was layered onto a 15-45% glycerol gradient (11 ml Beckman SW41, 125 mM KCl, 20 mM HEPES, 1.5 mM $MgCl_2$, 10% NP40, 10 mM 2-mercaptoethanol, 0.2 mM PMSF) made isotonic to LS- buffer (LS buffer without polyamines) and sedimented at 40,000 RPM for 90 minutes. Fractions (420 µl) were collected from the top. Protein and nucleic acid were separated by phenol/chloroform extraction and precipitation with acetone (protein) or ethanol (nucleic acid) (Stevens, 2000). Fractions corresponding to the supraspliceosomes were pooled from six velocity gradients run in a parallel fashion, diluted to ~ 8% glycerol with LS- buffer and incubated with 20 mg Y12 antibody which had been covalently attached to 1 g CnBr-sepharose (GE Biosciences) according to the manufacturer's instructions. After incubation with rotation for two hours at 4˚C, the sepharose matrix was washed with 200 ml LS- buffer

by gravity flow in a column and supraspliceosome material was eluted with 0.2 M glycine. Protein and nucleic acids were separated by phenol/chloroform extraction as described in section 2.2.7.

3.2.2 Purification of chicken supraspliceosome from DT40 cells

Six liters of SmD3-TAP DT40 cells (Chen et al., 2006) were grown in Dulbecco's modified Eagle media supplemented with 5% chicken serum and 2.5% Fetalplex (Gemini Bio-Products) to a density of 7.5×10^5 cells/ml for TAP purification. Cells were harvested by centrifugation (1000 x g for 5 minutes) and washed twice with ice cold PBS. Nuclear extracts preparation and TAP-tagged protein material for SmD3-TAP DT40 cells was affinity purified by the TAP procedure (Puig et al., 2001) as described in section 2.2.6 and 2.2.7. The TEV eluate was layered onto glycerol gradients and fractionated; fractions corresponding to the supraspliceosomes were phenol/chloroform extracted for proteins and nucleic acids analyses.

3.2.3 Mass spectrometry analysis

Pooled supraspliceosome protein fractions were separated by polyacrylamide gel electrophoresis and stained with Coomassie Blue G-250. Discrete gel slices were dissected from the top of the gel lane to the bottom and all regions were subjected to trypsin digestion. Mass spectrometry and database searching was performed as previously described in section 2.2.8.

3.2.4 Immunopurification of ZFR-associated components

3.2.4.1 Immunoprecipitation

Polyclonal antisera directed against the carboxyl-terminal 15 amino acids of KIAA0332 and NP_035897 (NCBI accession numbers) were produced by Genemed

Synthesis and the IgG fraction was partially purified by ammonium sulfate precipitation at 50% saturation. Antiserum or non-immune serum was incubated for one hour at 4˚C with the samples of interest prior to addition of 50 µl Protein-A agarose beads. This mixture was incubated one further hour with rotation at 4˚C prior to washing with 4 x 15 ml IPP150. Proteins and nucleic acids were released from the matrix by incubation in IPP150 at 100˚C for 5 minutes. The supernatant was collected and phenol/chloroform extracted as described above to harvest, separate and precipitate the nucleic acids and proteins.

3.2.4.2 Northern blot analysis

Nucleic acids were transferred to Brightstar membranes (Ambion) and hybridized with snRNA probes consisting of antisense chicken snRNAs transcribed with α^{32}P-GTP using T7 RNA polymerase (U5) or SP6 RNA polymerase (U1, U2, U4, U6 snRNAs) from plasmids containing cDNA versions of the chicken snRNAs.

3.2.4.3Western blot analysis

To determine if ZFR co-migrated with 200S particle in glycerol gradients, HeLa nuclear extracts were subjected to glycerol gradient sedimentation and resolved in 10% polyacrylamide gels (Graham et al., 2005), transferred to nitrocellulose membranes (Biorad) and blotted with anti-ZFR antiserum (1:1000 dilution). To determine if ZFR is specifically associated with spliceosomal proteins, hPrp43 antiserum was used to detect the presence of Prp43 (1:1000 dilution) (a gift of Cindy Will and Reinhard Lührmann). Anti-SR140 antiserum was used as a positive control (1:1000 dilution). The secondary antibody used was horseradish peroxidase-conjugated goat anti-rabbit IgG (Rockland; 1:10,000 dilution) and the signal was detected by enhanced chemiluminescence (Perkin Elmer).

3.3 RESULTS

3.3.1 Compositional analysis of the chicken supraspliceosome

Using the CLEP tagging procedure as described in Chapter 2, I tagged the SmD3 polypeptide in chicken DT40 cells with the TAP tag at its native genomic locus and purified the endogenous chicken pre-mRNA processing complexes from these cells.

These cells grow rapidly in suspension with an 8 hour doubling time, making them ideal for biochemical studies of this sort. For each experiment, six liters of SmD3-TAP-DT40 cells were harvested and the nuclear fraction was processed as described in 2.2.6 and 2.2.7. Affinity chromatography was performed according to the TAP procedure with the TEV protease digestion proceeding at 16˚C for 2 hours. The TEV eluate was sedimented through a glycerol gradient and the material corresponding to the supraspliceosomes, in the 200S region, was isolated. Associated proteins and nucleic acids are shown in Figure 3.1A and 3.1B, respectively. In Figure 3.2A and 3.2B, the proteins and RNA resulting from an identical affinity purification procedure performed using extracts from untagged DT40 cells are shown. The absence of proteins and snRNAs indicates that the purification is specific and the proteins identified by mass spectrometry are likely to be *bona fide* supraspliceosome components. Additional confidence is provided in that there is size-selection of these massive complexes as well as affinity chromatography.

3.3.1.1 RNA content of the supraspliceosome

RNAs corresponding to the supraspliceosome fractions from chicken cells are shown in Figure 3.1A. Identities of the RNAs were confirmed by northern blotting (data not shown). The presence of all five of the spliceosomal snRNAs in this material indicates that it contains a mixture of pre-mRNA splicing complexes in varying stages of

the splicing reaction as U4 snRNA has been shown to dissociate from the spliceosome before the first catalytic step in *vitro* reactions. Alternatively, it may be interpreted as a functional difference between the *in vitro* and *in vivo* nature of this analysis as compared to previous analyses and that the U4 snRNA is not removed, but only destabilized and maintained locally. The presence of all 5 snRNAs in roughly equivalent amounts also lends experimental evidence to the participation of the penta-snRNP in these functional complexes.

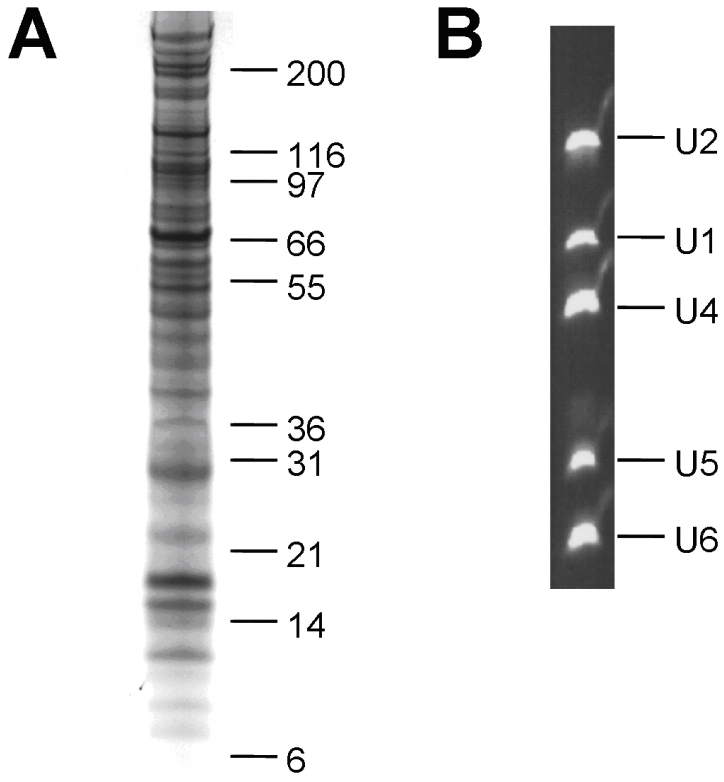

Figure 3.1 Chicken supraspliceosome-associated polypeptides and snRNAs.

Fractions corresponding to the CLEP-tag purified, glycerol gradient-sedimented chicken supraspliceosomes were separated into protein (A) and RNA (B) fractions and electrophoresed through SDS-PAGE (A) or urea-PAGE (B) gels and stained with coomassie blue (protein) or ethidium bromide (RNA). The identities of the snRNAs are indicated on the right of panel B. The entire gel lane from (A) was dissected and each gel slice was subjected to mass spectrometry protein identification. The proteins identified are reported under the Gg PS column in Tables 3.1-3.5, and 3.7.

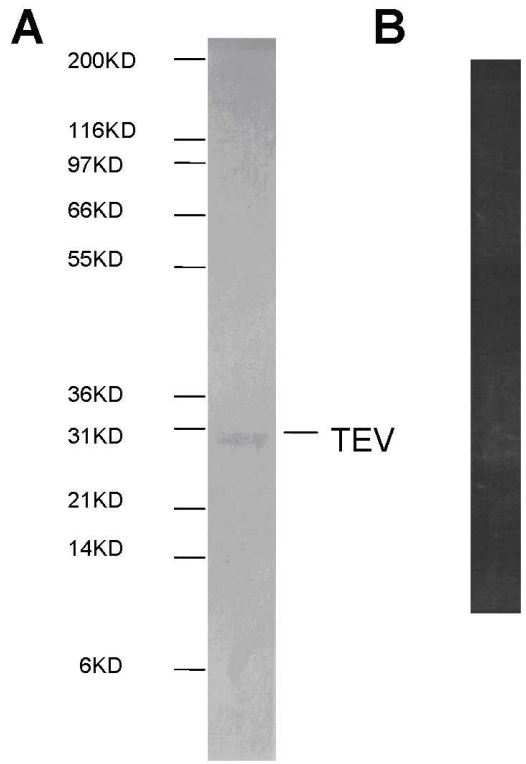

Figure 3.2 Mock-purified chicken supraspliceosomes.

(A) Proteins from a mock purification in which untagged DT40 extracts were subjected to the TAP purification procedure and the resulting> 200S material subjected to glycerol gradient sedimentation and stained with coomassie blue G-250. Note the presence of small amounts of residual TEV protease frequently seen distributed throughout the gradients. (B) Ethidium bromide-stained nucleic acids from the fractions.

3.3.1.2 Mass spectrometry analysis of supraspliceosome-associated polypeptides

Polyacrylamide gel material from the entire chicken supraspliceosome fraction lane was dissected and analyzed by tandem mass spectrometry (Davis and Lee, 1998; Moore et al., 2000; Moore et al., 2002). A wide variety of polypeptides were identified. We detected very little background contamination of factors unrelated to gene expression.

Remarkably, we detected 222 distinct polypeptides in the chicken supraspliceosomes (Table 3.1-3.5, and 3.7).

The polypeptides identified by mass spectrometry were validated by analyzing the numbers of peptides identified and percent-coverage for each protein (Appendix A). Although there is a distribution of coverage for the identified proteins, we note that some snRNP-associated proteins had less than 10% coverage. Differences in coverage may reflect differences in abundance, differences in the size spectrum of trypsin fragments or peculiarities in the mass spectrometry detection of a particular peptide. The low-coverage of some of the novel polypeptides that may reasonably be implied to function in pre-mRNA processing (i.e. contain RNA biding motifs) may reflect their association with a smaller subset of pre-mRNA than a general RNA binding protein such as an hnRNP. In the case of one novel factor, ZFR, the percent-coverage was low, 3.4% for chicken, but its association with the splicing machinery was verified independently in section 3.3.5.

In Tables 3.1-3.5, we categorize the identified polypeptides according to function. Polypeptides identified from the chicken supraspliceosome are briefly described (see text below).

Known snRNP-associated polypeptides

By mass spectrometry, we identified nearly all of the known pre-mRNA splicing snRNP-associated polypeptides (Table 3.1). Of the snRNP-associated proteins not present in the mass spectrometry results, their inability to be detected correlates with the presence of an extramolar hnRNP protein of a similar molecular weight. The CLEP tagging procedure was sensitive enough to detect the presence of two minor AT-AC spliceosome components, the U11/U12-65K and U11/U12-48K polypeptides (Will et al., 2004) in the chicken fractions. The ability to detect all of the Sm proteins, but not all of the Lsm proteins may reflect the difficulty in detecting all of these proteins in splicing

complexes as shown previously (Gottschalk et al., 1999; Stevens and Abelson, 1999), the five-fold abundance differences of the two classes of proteins or perhaps due to the Lsm proteins leaving the spliceosome during the process of pre-mRNA splicing (Chan et al., 2003).

snRNP biogenesis factors

Previous spliceosome purifications did not yield polypeptides known to be involved in snRNP biogenesis. We note in Table 3.1 that there are three present in the chicken supraspliceosomes (SIP1, SMNrp30 and Coilin). These are factors which are involved in the *de novo* assembly of snRNPs and are contained in Cajal bodies (CBs), a site of enriched pre-mRNA splicing factors (Gall et al., 1999). We hypothesize that we may be purifying a small subset of snRNPs in the process of being re-targeted to the CBs.

Table 3.1 Comparisons of snRNPs and snRNP biogenesis factors between supraspliceosomes purified from human or chicken cells with those purified from three *in vitro* assembled spliceosome experiments.

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
U1 snRNP							
ENSG00000104852	SNRP70	U1-70K	•	•	•	•	•
ENSG00000077312	SNRPA	U1A	•		•	•	•
ENSG00000124562	SNRPC	U1-C				•	•
U2 snRNP							
ENSGALG00000008038	SF3B1	SF3b155	•	•		•	•
ENSG00000087365	SF3B2	SF3b145	•	•	•	•	•
ENSGALG00000002531	SF3B3	SF3b130	•	•	•	•	•
ENSGALG00000000581	DDX42	SF3b125	•	•			
ENSGALG00000007937	SF3A1	SF3a120	•	•		•	•
ENSGALG00000021679	SF3A2	SF3a66	•	•	•	•	•
ENSGALG00000001540	SF3A3	SF3a60	•	•	•	•	•
ENSGALG00000013352	SF3B4	SF3b49	•		•	•	•
ENSGALG00000008729	SNRPB2	U2B"	•	•	•	•	•
ENSGALG00000007170	SNRPA1	U2A'	•	•	•	•	•
ENSGALG00000016501	-	SF3b14	•			•	
ENSGALG00000020000	SF3B5	SF3b10	•			•	
U2-snRNP associated							
ENSGALG00000014395	DHX15	PRP43/DDX15	•	•		•	•
ENSGALG00000002612	-	SR140	•	•		•	•
ENSGALG00000006332	RBM17	SPF45	•		•	•	•
ENSGALG00000008561	SMNDC1	SPF30	•		•	•	•
ENSGALG00000003824	CHERP	CHERP	•		•	•	
U5, U4/U6 & U4/U6•U5 snRNP							
ENSGALG00000002943	PRPF8	U5-220K	•	•		•	•
ENSGALG00000003477	ASCC3L1	U5-200K	•	•		•	•
ENSGALG00000000988	EFTUD2	U5-116K	•	•		•	•
ENSG00000175467	SART1	U4/U5•U5-110K	•	•		•	•
ENSGALG00000006001	C20ORF14	U5-102K	•	•		•	•
ENSG00000174243	DDX23	U5-100K	•	•	•	•	•
ENSGALG00000000465	PRPF3	U4/U6-90K	•	•	•	•	•
ENSG00000168883	USP39	U4/U6•U5-65K	•	•		•	•

Table 3.1 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSG00000105618	PRPF31	U4/U6•U5-61K	•	•		•	•
ENSGALG00000008857	PRPF4	U4/U6-60K	•	•	•	•	•
ENSGALG00000004874	PPIH	USA-CYP	•				•
ENSGALG00000000615	WDR57	U5-40K	•	•	•		
ENSGALG00000011931	NHP2L1	U4/U6•U5-15.5k	•			•	•
ENSGALG00000017396	TXNL4A	U5-15K	•				•
AT/AC							
ENSGALG00000005162	RNPC3	U11/U12-65K	•				
ENSGALG00000013005	C6ORF151	U11/U12-48K	•				
Sm/LSM							
ENSGALG00000007250	SNRPB	SmB/B'	•	•	•	•	•
ENSGALG00000011842	SNRPD1	SmD1	•	•		•	•
ENSG00000125743	SNRPD2	SmD2	•	•		•	•
ENSGALG00000006596	SNRPD3	SmD3	•	•		•	•
ENSGALG00000000137	SNRPE	SmE	•	•		•	•
ENSGALG00000011409	SNRPF	SmF	•	•		•	•
ENSG00000143977	SNRPG	SmG	•	•		•	•
ENSG00000111987	LSM2	LSM2	•			•	•
ENSG00000170860	LSM3	LSM3				•	•
ENSGALG00000003385	LSM4	LSM4	•			•	•
ENSG00000106355	LSM5	LSM5					
ENSGALG00000009985	LSM6	LSM6	•	•		•	•
ENSG00000130332	LSM7	LSM7				•	•
ENSGALP00000014820	LSM8	LSM8	•	•		•	
snRNP biogenesis							
ENSGALG00000010154	SIP1	SIP1	•				
ENSG00000119953	SMNDC1	SMNrp30	•				
ENSGALG00000003158	COIL	Coilin	•				

[a] Data available at http://www.ensembl.org; [b] HUGO Gene Nomenclature Committee designation, a dash represents a polypeptide which has not yet been assigned a systematic name; [c] Common name; [d] *Gallus gallus* supraspliceosome; [e] *Homo sapiens* (HeLa) supraspliceosome; [f] Neubauer *et al*. spliceosome data (Neubauer et al., 1998); [g] Rappsilber *et al*. spliceosome data (Rappsilber et al., 2002); [h] Zhou *et al*. spliceosome data (Zhou et al., 2002). [f,g,h] were compiled from Jurica and Moore (Jurica and Moore, 2003).

Known spliceosome associated proteins (SAPs)

In Table 3.2, we outline the 41 SAPs identified from the chicken polyspliceosome. Included are pre-mRNA interacting factors such as U2AF (Bennett et al., 1992; Kramer and Utans, 1991), PTB (Gil et al., 1991; Patton et al., 1991), and the cap binding complex proteins (Lewis et al., 1996). We detected the majority of PRP19-complex (NTC) related components as well, including homologues of Prp19p (Cheng et al., 1993; Russell et al., 2000), Syf1p (Russell et al., 2000), Syf2p (Russell et al., 2000), Syf3/Clf1p (Chung et al., 1999; Russell et al., 2000)), Isy1p (Dix et al., 1999), SKIP/Prp45p (Albers et al., 2003; Figueroa and Hayman, 2004) and CDC5/Cef1p (Ben-Yehuda et al., 2000; Tsai et al., 1999). Interestingly, we detected a number of polypeptides which in yeast are snRNP associated, but have not been identified in purified metazoan snRNPs. Among these polypeptides are putative orthologues of yeast Prp38p (Blanton et al., 1992), Prp39p (Lockhart and Rymond, 1994), Prp40p (Kao and Siliciano, 1996), Aar2p (Gottschalk et al., 2001) and Luc7p (Fortes et al., 1999). Other polypeptides exclusively contained in the purified chicken supraspliceosome include, NONO/p54nrb, PNN/Pinin and CWC22, which have been previously implicated in splicing (Kameoka et al., 2004; Ohi et al., 2002; Sakashita et al., 2004; Wang et al., 2002)).

Table 3.2 Comparisons of spliceosome-associated proteins between supraspliceosomes purified from human or chicken cells with those purified from three *in vitro* assembled spliceosome experiments.

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
SAPs							
ENSGALG00000002514	SFPQ	PSF	•	•		•	
ENSGALG00000002060	FUS	TLS/FUS	•			•	•
ENSGALG00000012468	PRPF39	PRP39	•				
ENSGALG00000010501	SNW1	SKIP/PRP45	•	•	•	•	•
ENSGALG00000016704	CDC5L	CDC5	•	•	•		•
ENSGALG00000005012	RAB43	ISY1	•			•	•
ENSGALG00000008429	CRNKL1	CRN1	•	•	•	•	•
ENSGALG00000009257	PRLG1	Prp46/PRL1	•			•	•
ENSGALG00000015061	CDC40	CDC40/PRP17	•			•	•
ENSGALG00000002002	BCAS2	SPF27	•	•			
ENSGALG00000013919	PRPF19	PRP19	•	•	•	•	•
ENSGALG00000010627	PRPF38A	PRP38	•	•			
ENSGALG00000005507	NONO	p54nrb	•	•			
ENSGALG00000004555	RBM22	ECM2/RBM22	•			•	•
ENSGALG00000001247	SYF2	SYF2	•			•	•
ENSGALG00000000726	ELAVL1	ELAV/Hu	•	•		•	•
ENSGALG00000009001	-	CWC22	•				
ENSGALG00000008149	EWSR1	EWSR1 (RBP)	•	•		•	
ENSGALG00000004705	BUD31	BUD31	•			•	•
ENSG00000076924	XAB2	SYF1	•	•		•	•
ENSGALG00000001962	PTBP1	PTB	•	•		•	
ENSGALG00000011857	LUC7L2	LUC7/CROP	•				•
ENSG00000196504	PRPF40A	PRP40	•	•			•
ENSGALG00000010167	PNN	Pinin	•	•			
ENSGALG00000012813	PRPF4B	Prp4K	•	•		•	•
ENSGALG00000000833	IK	RED	•	•		•	•
ENSGALG00000001500	-	SLU7	•			•	•
ENSG00000063244	U2AF2	U2AF65	•	•	•	•	•
ENSGALG00000016198	U2AF1	U2AF35	•	•	•	•	•
ENSG00000168066	SF1	SF1			•	•	•
ENSGALG00000001034	C20ORF4	AAR2	•				
ENSGALG00000005525	SFRS1	SF2/ASF	•	•		•	•
ENSG00000102241	HTATSF1	TAT-SF1					•
ENSGALG00000002087	NCBP1	CBC80	•	•		•	•
ENSGALG00000006843	NCBP2	CBC20	•	•		•	•

Table 3.2 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSG00000087087	-	ASR2B				•	•
ENSG00000100296	THOC5	KIAA0983	•	•		•	•
ENSG00000159086	C21ORF66	C21ORF66				•	•
ENSG00000126803	HSPA2	HSP70-2	•	•		•	•
ENSGALG00000006512	HSPA8	HSP71	•	•		•	•
ENSGALG00000009838	AQR	Aquarius	•	•		•	•
ENSGALG00000002014	SMU1	SMU1	•	•		•	•
ENSG00000163510	-	CWC22				•	•
ENSGALG00000005623	TFIP11	SPP382	•	•		•	•
ENSG00000137656	-	CWC26				•	•
ENSG00000126698	DNAJC8	SPF31			•	•	•
ENSG00000105705	SF4	SF4			•	•	
ENSG00000113649	TCERG1	CA150		•	•	•	•
ENSGALG00000004626	RBM5	E1B-AP5	•	•		•	
ENSG00000100056	DGCR14	DGCR14				•	•
ENSG00000105298	C19ORF29	C19ORF29				•	•
ENSG00000109536	FRG1	FRG1				•	
ENSG00000171824	EXOSC10	RRP6				•	
ENSG00000160799	CCDC12	CCDC12/CWF18				•	
ENSGALG00000011678	DNAJC13	DnaJ	•	•		•	
ENSG00000100813	ACIN1	Acinus		•		•	•
ENSG00000131051	RNPC2	HCC		•		•	•
ENSG00000084463	WBP11	WBP11		•		•	
ENSG00000196419	XRCC6	Ku70		•		•	

Supraspliceosome-associated hnRNPs

Polypeptides termed hnRNPs are those which are known to interact with hnRNA and are highly abundant nuclear proteins. We detected virtually all of the known hnRNP proteins in chicken cells, as well as other hnRNP-like proteins present in genome databases (Table 3.3).

RNA helicase-like proteins

We identified a large number of DExH/D proteins in the purifications (Table 3.3). In addition to the RNA helicase-like polypeptides known to function in pre-mRNA splicing, such as DDX15/Prp43p (Arenas and Abelson, 1997; Fouraux et al., 2002; Martin et al., 2002), DHX8/Prp22p (Company et al., 1991; Schwer and Gross, 1998; Wagner et al., 1998), DDX46/Prp5p (Dalbadie-McFarland and Abelson, 1990; O'Day et al., 1996; Will et al., 2002), DDX5/p68 and DDX17/p72 (Lin et al., 2005), UAP56/Sub2p (Kistler and Guthrie, 2001; Libri et al., 2001; Zhang and Green, 2001), and the snRNP-associated helicases U5-200K/Brr2p (Lauber et al., 1996; Noble and Guthrie, 1996; Raghunathan and Guthrie, 1998) and U5-100K/Prp28p (Strauss and Guthrie, 1991; Strauss and Guthrie, 1994; Teigelkamp et al., 1997), we note a number of helicase-like proteins not previously implicated in pre-mRNA splicing contained in purified chicken supraspliceosomes. These include 13 additional polypeptides with sequence motifs indicative of DEAD, DEAH or Ski2p-like helicase family members. Although we do not have evidence for these proteins functioning in pre-mRNA splicing, the complete absence of DNA helicases in our preparations indicates a specificity which implies their function in RNA Pol II transcript processing.

SR proteins

Many splicing factors are rich in arginine and serine residue including long stretches of alternating dipeptides termed SR domains. These factors function at multiple steps in the pre-mRNA splicing pathway (Bourgeois et al., 2004). We detected a number of these family members in our purifications from chicken cells (Table 3.3). Although there is not a complete set of known SR-proteins present, it is possible to conclude that, due to the means by which these complexes were purified and analyzed,

these SR proteins represent the major SR proteins functioning in these cells, and that the others function in the splicing of a smaller subset of pre-mRNAs.

Other RNA binding proteins

There were 18 polypeptides identified in our purified polyspliceosome material possessing sequence homology to polypeptides predicted to interact with RNA by virtue of RNA Recognition Motifs (RRM), double stranded RNA binding domains (dsRBD) and others (Table 3.3). Some were identified previously, such as the ELAV/Hu protein involved in binding to AU-rich elements (Myer et al., 1997), the dsRBD-motif-containing NFAT45 and NFAT90, previously shown to exist in large nuclear complexes (Raitskin et al., 2001) and likely functioning in RNA Pol II transcript metabolism. In Figure 3.3 we present a graphical representation of the various RNA binding proteins novel to our study and highlight their detectable sequence motifs.

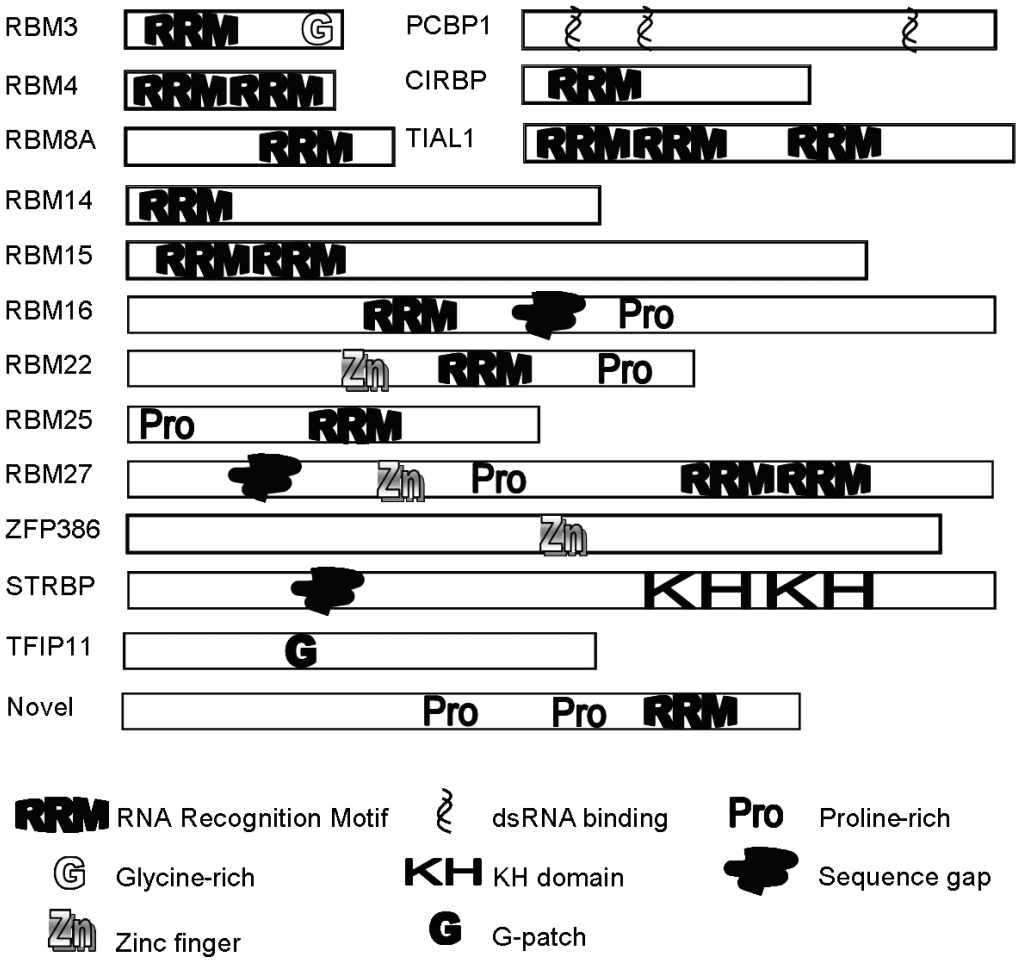

Figure 3.3 Novel spliceosome-associated polypeptides with predicted RNA binding motifs.

Polypeptides from Table 3.3 with no known function in the pre-mRNA processing pathway are shown with graphical representations of the various RNA interaction or other noted motifs listed at the bottom of the Figure.

Non-spliceosomal pre-mRNA processing factors

In Table 3.1, we note the presence of 5' end binding (CBC80 and CBC20) proteins. In Table 3.3, the 3' end processing factors (CSPF and CSTF), a comprehensive set of proteins shown to be involved in mRNA export including the TREX complex (THO1/HPR1, THO2, THO3/TEX1, UAP56 and ALY) (Strasser et al., 2002) and export factors such as GLE2 (Blevins et al., 2003) and GLC7 (Gilbert and Guthrie, 2004) are reported. We also note that a single component of the nonsense mediated decay (NMD) pathway, UPF1, (Mitchell and Tollervey, 2003) was identified in the chicken material. As NMD is likely to be active only in a very small subset of pre-mRNA processing complexes, we were surprised to see even a single polypeptide from this class of proteins.

Table 3.3 Polypeptides predicted to interact with the pre-mRNA, mRNA or the spliceosome and comparisons of those identified in the supraspliceosomes with those of spliceosomes formed *in vitro*.

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
RNA Helicase-like							
ENSGALG00000016461	DDX1	DDX1	•	•			
ENSGALG00000016231	DDX3X	DDX3	•	•			•
ENSGALG00000003532	DDX5	DDX5/p68	•		•	•	•
ENSGALG00000012247	DDX17	DDX17/p72	•	•		•	•
ENSGALG00000012147	DDX18	DDX18	•				
ENSG00000102786	DDX26	DDX26/HDB				•	
ENSGALG00000006974	DDX27	DDX27	•				
ENSGALG00000003030	DDX41	DDX41/ABSTRAKT	•			•	•
ENSG00000123136	DDX39	DDX39		•			
ENSG00000145833	DDX46	DDX46/PRP5		•		•	•
ENSGALG00000008530	DDX48	DDX48	•	•		•	•
ENSGALG00000004144	DDX50	DDX50/Gu-ß	•	•			
ENSGALG00000001186	DHX8	DHX8/PRP22	•	•		•	•
ENSG00000135829	DHX9	DHX9/HELICASEA	•	•		•	•
ENSG00000137333	DHX16	DHX16/PRP2				•	•
ENSGALG00000005027	DHX30	DHX30	•	•			
ENSGALG00000003658	DHX35	DHX35	•				
ENSG00000174953	DHX36	DHX36		•			
ENSGALG00000014709	SKIV2L2	SKIV2L2	•	•		•	•
ENSG00000198563	BAT1	UAP56	•	•			•
hnRNP							
ENSGALG00000006160	HNRNPA0	hnRNPA0	•	•			
ENSGALG00000011036	HNRNPA2B1	hnRNPA2/B1	•	•	•	•	•
ENSG00000135486	HNRPA1	hnRNPA1	•	•	•	•	•
ENSGALG00000009250	HNRNPA3	hnRNPA3	•	•		•	•
ENSGALG00000014381	HNRNPAB	hnRNPAB	•	•			
ENSG00000092199	HNRPC	hnRNPC1/C2	•	•	•	•	•
ENSG00000138668	HNRNPD	hnRNPD0/AUF1	•	•			
ENSGALG00000011184	HNRNPDL	hnRNPD0	•	•		•	•
ENSG00000169813	HNRPF	hnRNPF				•	
ENSGALG00000006457	RBMX	hnRNPG	•	•		•	
ENSGALG00000005955	HNRNPH1	hnRNPH1	•	•		•	
ENSGALG00000003947	HNRNPH3	hnRNPH3	•	•			
ENSGALG00000012591	HNRNPK	hnRNPK	•	•		•	•

Table 3.3 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSG00000104824	HNRNPL	hnRNPL	•	•		•	•
ENSGALG00000000377	HNRNPM	hnRNPM	•	•	•	•	
ENSGALG00000015830	SYNCRIP	hnRNPQ	•	•		•	•
ENSGALG00000000814	HNRNPR	hnRNPR	•	•		•	•
ENSGALG00000010671	HNRNPU	hnRNPU	•	•		•	
ENSGALG00000018665	-	hnRNP novel	•				
ENSG00000126457	HRMT1L2	HRMT1L2	•				
SR							
ENSG00000133226	SRRM1	SRm160				•	•
ENSG00000167978	SRRM2	SRm300				•	•
ENSGALG00000005525	SFRS1	SF2p33	•				
ENSG00000161547	SFRS2	SC35				•	•
ENSGALG00000000533	SFRS2	SFRS3SRp20	•	•		•	•
ENSG00000116350	SFRS4	SRp75					•
ENSGALG00000009484	SFRS5	SRp40	•				•
ENSGALG00000000990	SFRS6	SRp55	•	•		•	•
ENSGALG00000013825	SFRS7	9G8	•	•		•	•
ENSGALG00000002487	SFRS8	SFRS8	•				
ENSG00000111786	SFRS9	SRp30		•		•	•
ENSGALG00000006531	SFRS10	SFRS10	•	•			•
ENSG00000116754	SFRS11	SRp54				•	•
ENSG00000153914	SFRS12	SFRS12		•			
ENSGALG00000004133	FUSIP1	SRrp35	•	•			
RBP							
ENSG00000102317	RBM3	RBM3		•			
ENSGALG00000018992	RBM4B	RBM4B/Lark	•				
ENSGALG00000004626	RBM5	RBM5	•			•	
ENSGALG00000007017	RBM7	RBM7	•			•	
ENSG00000131795	RBM8A	RBM8B		•			
ENSGALG00000013411	RBM14	RBM14	•	•			
ENSG00000162775	RBM15	RBM15		•		•	•
ENSGALG00000002372	RBM15B	RBM15B	•	•			
ENSG00000197676	RBM16	RBM16		•			
ENSGALG00000004555	RBM22	RBM22	•				
ENSG00000119707	RBM25	RBM25		•			
ENSG00000091009	RBM27	RBM27		•			

Table 3.3 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSGALG00000011038	CBX3	RNPS1	•				•
ENSG00000033030	ZCCHC8	ZFP8		•		•	
ENSGALG00000006113	ZNF326	ZFP326	•				
ENSG00000179950	-	PUF60		•		•	•
ENSG00000197381	ADARB1	ADAR1		•			
ENSG00000160710	ADAR	ADAR2		•			
ENSGALG00000010952	-	Requiem	•				
ENSGALG00000011570	ILF2	NFAT45	•	•		•	•
ENSG00000129351	ILF3	NFAT90	•	•		•	•
ENSG00000169564	PCBP1	PolyrCBP	•				
ENSGALG00000012225	CIRBP	CIRBP	•	•			
ENSG00000056097	ZFR	ZFR	•	•			
ENSGALG00000009427	TIAL1	TIA1	•				
ENSGALG00000001187	STRBP	STRBP	•				
ENSGALG00000005623	TFIP11	Tuftillin-IP	•	•			
ENSG00000136231	IGF2BP3	IMP3	•	•		•	•
ENSG00000060138	CSDA	CSDA				•	•
ENSG00000121774	KHDRBS1	SAM68		•			
ENSG00000126254	-	Novel RRM		•			
ENSG00000132773	TOE1	TOE1		•			
ENSG00000142864	SERBP1	SERBP1		•			
Export/transcription/NMD							
ENSGALG00000014915	THOC1	THO1/HPR1	•	•			
ENSGALG00000008507	THOC2	THO2	•	•		•	•
ENSG00000051596	THOC3	THO3/TEX1	•	•			
ENSGALG00000007237	THOC4	THO4/ALY	•	•		•	•
ENSGALG00000004571	PPP1CA	GLC7/PPP1CA	•				
ENSGALG00000002569	RAN	Ran	•	•			
ENSG00000119392	GLE1L	GLE1L		•			
ENSGALG00000007653	RAE1	GLE2/RAE1	•	•			
ENSGALG00000003220	RENT1	UPF1/RENT	•				
ENSG00000131795	RBM8A	Y14/RBM8A	•	•			•
ENSGALG00000002144	THRAP3	TRAP150	•	•			
ENSG00000172660	TAF15	TAF15/RBP56	•				
ENSG00000162231	NXF1	TAP		•			
ENSG00000065978	YBX1	YBX1		•		•	•

Table 3.3 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSGALG00000010689	MAGOH	Mago nashi	•	•			•
3' end proc.							
ENSG00000071894	CPSF1	CPSF1				•	
ENSGALG00000010783	CPSF2	CPSF2	•				
ENSGALG00000016424	CPSF3	CPSF3	•	•			
ENSGALG00000004714	CPSF4	CPSF4	•				
ENSGALG00000003084	CPSF5	CPSF5		•		•	•
ENSG00000111605	CPSF6	CPSF6				•	•
ENSGALG00000011685	CSTF3	CSTF-77	•	•			
ENSGALG00000013943	FIP1L1	FIP1	•				
ENSG00000172239	PAIP1	PAIP1		•			
ENSG00000100836	PABPN1	PABPN1		•	•	•	•
ENSGALG00000003800	PABPC4	PABPC4		•			

Nuclear matrix and filament proteins

Recent data from several laboratories indicates functional interaction between the structural proteins of the nuclear matrix and the gene expression machinery (de Lanerolle et al., 2005). We detected a number of such proteins including actin, spectrin, matrin3, numatrin, and lamin B (Table 3.4). Although we cannot determine the functional relevance of the association of structural proteins with these endogenously formed pre-mRNA splicing complexes, we note that a number of hnRNPs and other known splicing factors such as Prp19p (Gotzmann et al., 2000) were initially termed nuclear matrix-associated proteins indicating an intimate relationship between the pre-mRNA processing machinery and the nuclear matrix. Indeed it is an attractive hypothesis that pre-mRNA and mRNA are trafficked to the nuclear pore by means of the nuclear matrix.

Polypeptides novel to pre-mRNA splicing- SWI/SNF proteins and associated factors

A recent report from Muchardt and colleagues has demonstrated the association of the SWI/SNF associated factor Brahma/SMARCA4 (Brm) with the splicing apparatus and its presence favors the inclusion of alternatively spliced exons (Batsche et al., 2006). The splicing-related Prp4-kinase, which is present in both the human and chicken supraspliceosomes, has been reported to phosphorylate both Brm and the splicing factor U5-102K/hPrp6 (Dellaire et al., 2002) providing further evidence that Brahma is functioning in Pol II transcript maturation. In the purified chicken supraspliceosomes, Brahma/SMARCA4, and a number of other SWI/SNF-related polypeptides were identified (Table 3.4), all with high degrees of confidence given the depth of the peptide identification (Appendix A). As this purification yielded no mass spectrometry data suggesting any chromatin-associated factors such as histone proteins, the DNA replication machinery or other DNA binding proteins, we have no reason to believe that these polypeptides are not specifically associating with the pre-mRNA processing complexes. Further studies will illuminate the specific function of these polypeptides in pre-mRNA processing.

Nuclear pore complex proteins

In the chicken supraspliceosomes, we detected a small set of NUPs (Table 3.4). It will be interesting to see if these are functional interactions between maturing pre-mRNA processing complexes and the nuclear pore, or if these polypeptides are promiscuously interacting with the supraspliceosomes.

Cyclophillins

In addition to the known snRNP-associated USA-CYP (Horowitz et al., 1997; Teigelkamp et al., 1998), we detected 5 additional potential disulfide isomerases co-

purifying with spliceosomes from chicken (Table 3.4). These proteins are important for folding and structural rearrangement.

Table 3.4 Structural, nucleoporin, and cyclophillin proteins present in the supraspliceosomes and *in vitro* assembled splicing complexes

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
Structural							
ENSGALG00000012533	MYH9	Myosin	•	•			
ENSGALG00000009126	TTN	Titin	•				
ENSGALG00000001381	ACTG1	Actin	•	•			
ENSGALG00000002478	MATR3	MATRIN3	•	•		•	•
ENSGALG00000008677	VIM	Vimentin	•	•			
ENSG00000117245	KIF17	Kinesin KIF17	•				
ENSGALG00000002197	NPM1	NUMATRIN	•	•		•	
ENSGALG00000014692	LMNB1	Lamin B	•	•			
ENSGALG00000013505	SYNE1	NuSpectrin	•	•			
ENSG00000140259	MFAP1	MFAP1				•	•
Chromatin modification							
ENSGALG00000000360	ARID1A	ARID1A-SWI/SNF	•				
ENSGALG00000013683	ARID1B	ARID1B-SWI/SNF	•				
ENSGALG00000010164	SMARCA2	SMARCA2	•				
ENSG00000127616	SMARCA4	Brahma/SMARCA4	•				
ENSGALG00000009913	SMARCA5	SMARCA5	•				
ENSGALG00000005983	SMARCB1	SMARCB1	•				
ENSGALG00000005048	SMARCC2	BRG1-SWI/SNF	•				
ENSGALG00000005048	SMARCC1	SMARCC1	•				
ENSGALP00000010010	SMARCD1	SMARCD1	•				
ENSGALG00000000363	SMARCD2	SMARCD2	•				
ENSGALG00000002100	SMARCE1	SMARCE1	•				
Nucleoporins							
ENSGALG00000005714	PKD1	PKD1 (NUP assoc)	•				
ENSGALG00000003830	NUP214	NUP214	•	•			
ENSGALG00000012720	NUP153	NUP153	•				
ENSGALG00000005078	NUP210	NUP210	•	•			
ENSG00000102900	NUP93	NUP93		•			
ENSG00000108559	NUP88	NUP88		•			
ENSG00000111581	NUP107	NUP107		•			
ENSG00000110713	NUP98	NUP98		•			
ENSG00000138750	NUP54	NUP54		•			
ENSG00000163002	NUP35	NUP35		•			
ENSG00000069248	NUP133	NUP133		•			

Table 3.4 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
ENSG00000155561	NUP205	NUP205		•			
Cyclophillins							
ENSGALG00000013383	PPIE	CYP-E	•			•	•
ENSGALG00000004874	PPIH	USA-CYP	•				
ENSGALG00000014747	SDCCAG10	CYP16	•			•	
ENSG00000137168	PPIL1	PPIL1/CWF27	•			•	•
ENSG00000100023	PPIL2	PPIL2/CYP60	•			•	•
ENSG00000115934	PPIL3	PPIL3B				•	•
ENSG00000113593	PPWD1	PPWD1	•			•	•

3.3.2 Compositional comparison of chicken versus human supraspliceosome

For purification of the human supraspliceosomes, we gently sonicated nuclei in a buffered low-salt solution. After a short, low-speed centrifugation, the resulting nuclear extract was sedimented through a glycerol gradient and the material from a region corresponding to a sedimentation coefficient >200S was pooled (Figure 3.4A, 3.4B). We did not observe a peak sedimenting at a certain gradient fraction due to the fact that the lnRNP particle is proposed to have one to eight repeating spliceosomal substructures. This material was incubated with anti-Sm epitope-reactive monoclonal antibody Y12 covalently attached to Sepharose beads to bind Sm-protein-containing material to the affinity resin. Remarkably, this treatment nearly quantitatively retained the detectable material from this region of the glycerol gradient, even after extensive washing indicating that the majority of the nuclear contents of this size, perhaps not surprisingly, are Sm-antigen-containing complexes. The lack of polyribosomes in the rapidly sedimenting material (as judged by the absence of 5S and 5.8S rRNAs and ribosomal proteins by mass

spectrometry) indicates that nuclei we prepare are not contaminated with cytoplasm. Preparatively-purified material was separated by SDS-PAGE gels of two compositions to better resolve the polypeptides (Figure 3.4C). To demonstrate the specificity of this purification, gradient-separated supraspliceosomes were subjected to affinity chromatography using identical beads and identical washing and elution conditions, but lacking the Y12 antibody. In Figure 3.5A and 3.5B, we show that from mock purification, there is not detectable coomassie-stained material in the resulting protein gel and no snRNAs are present.

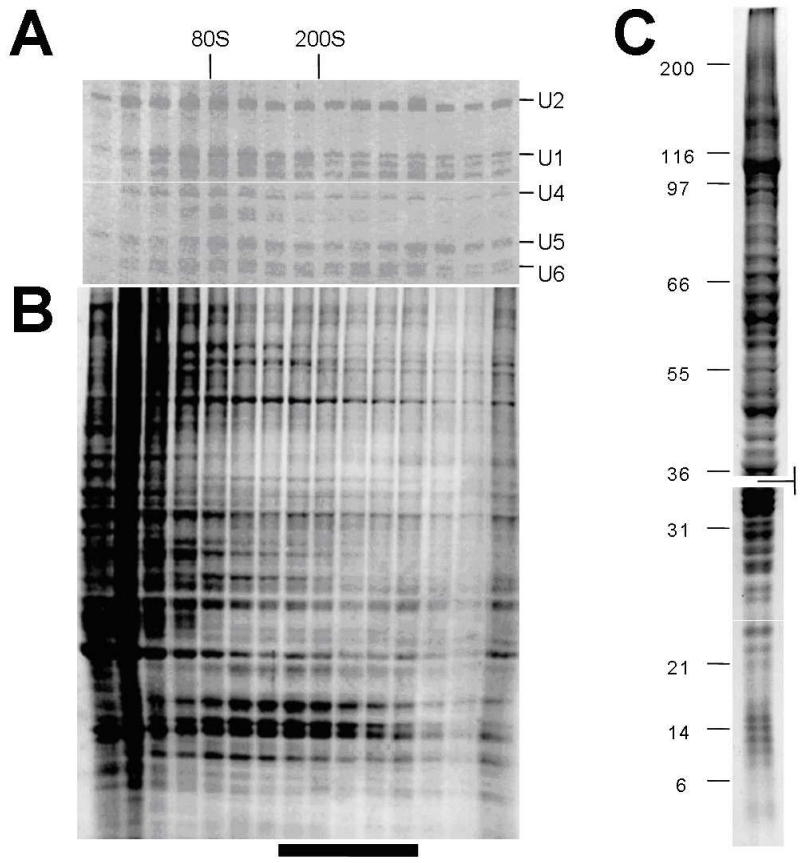

Figure 3.4 Human supraspliceosome-associated polypeptides and snRNAs.

RNA (A) and protein (B) were extracted from preparative glycerol gradient fractions and electrophoretically resolved through urea-PAGE (A) or SDS-PAGE (B) gels stained with silver (nucleic acid) or coomassie blue (protein). Bar below B represents the fractions of the material pooled for immunopurification with Y12 antibody. (C) Affinity-purified supraspliceosomal proteins run under two SDS-PAGE conditions to resolve either large or small polypeptides. Gels were aligned to show all polypeptides in the affinity-purified fractions and are delineated by the marking between them. The entire gel lanes shown from the two gels in (C) were dissected and each gel slice was subjected to mass spectrometry protein identification. The proteins identified are reported under the Hs PS column in Tables 3.1-3.5 and 3.7.

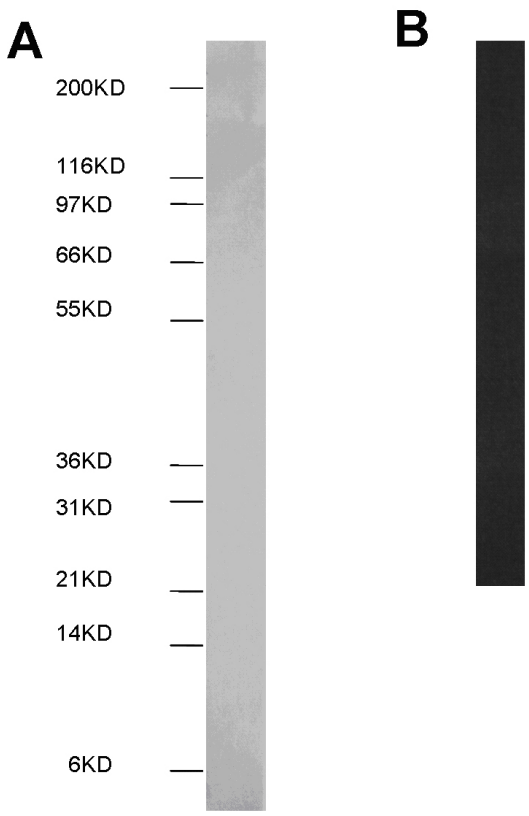

Figure 3.5 Mock-purified HeLa supraspliceosomes.

(A) Proteins from HeLa mock-purified polyspliceosome stained with coomassie blue G-250. (B) Ethidium bromide-stained nucleic acids from the fraction.

By mass spectrometry we detected 177 distinct polypeptides in the HeLa supraspliceosome, however, the coverage of the major spliceosomal snRNP proteins was more complete for the chicken supraspliceosomes. Polypeptides identified specifically from chicken supraspliceosome but not from HeLa supraspliceosome are summarized in Table 3.5.

No AT-AC specific splicing components were detected in the human complex, suggesting CLEP tagging and purification procedure is more sensitive than conventional antibody affinity chromatography. We did not detect any snRNP biogenesis factors in the HeLa supraspliceosomes, but we note that the chicken cells were processed fresh, while the HeLa cells were obtained commercially and remained at 4˚C during shipment overnight. This may disfavor the purification of these or even other factors (e.g. cyclophillins and chromatin remodeling proteins) in the case of the HeLa supraspliceosomes. A substantial set of nucleoporins (NUPs) were present in the purified supraspliceosomes complexes from human cells. NP40 was absent during the purification of the human complexes, which likely maintained the integrity of hydrophobic interactions which may stabilize the export complexes with the NUPs.

Table 3.5 Polypeptides identified in chicken supraspliceosomes but not in human supraspliceosomes

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]
U2 snRNP				
ENSGALG00000013352	SF3B4	SF3b49	•	
ENSGALG00000016501	-	SF3b14	•	
ENSGALG00000020000	SF3B5	SF3b10	•	
U2-snRNP associated				
ENSGALG00000006332	RBM17	SPF45	•	
ENSGALG00000008561	SMNDC1	SPF30	•	
ENSGALG00000003824	CHERP	CHERP	•	
U5, U4/U6 & U4/U6•U5 snRNP				
ENSGALG00000004874	PPIH	USA-CYP	•	
ENSGALG00000011931	NHP2L1	U4/U6•U5-15.5k	•	
ENSGALG00000017396	TXNL4A	U5-15K	•	
AT/AC				
ENSGALG00000005162	RNPC3	U11/U12-65K	•	
ENSGALG00000013005	C6ORF151	U11/U12-48K	•	
Sm/LSM				
ENSG00000111987	LSM2	LSM2	•	
ENSGALG00000003385	LSM4	LSM4	•	
snRNP biogenesis				
ENSGALG00000010154	SIP1	SIP1	•	
ENSG00000119953	SMNDC1	SMNrp30	•	
ENSGALG00000003158	COIL	Coilin	•	
SAPs				
ENSGALG00000002060	FUS	TLS/FUS	•	
ENSGALG00000012468	PRPF39	PRP39	•	
ENSGALG00000005012	RAB43	ISY1	•	
ENSGALG00000009257	PRLG1	Prp46/PRL1	•	
ENSGALG00000015061	CDC40	CDC40/PRP17	•	
ENSGALG00000004555	RBM22	ECM2/RBM22	•	
ENSGALG00000001247	SYF2	SYF2	•	
ENSGALG00000009001	-	CWC22	•	
ENSGALG00000004705	BUD31	BUD31	•	
ENSGALG00000011857	LUC7L2	LUC7/CROP	•	
ENSGALG00000001500	-	SLU7	•	
ENSGALG00000001034	C20ORF4	AAR2	•	
RNA Helicase-like				
ENSGALG00000003532	DDX5	DDX5/p68	•	

Table 3.5 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]
ENSGALG00000012147	DDX18	DDX18	•	
ENSGALG00000006974	DDX27	DDX27	•	
ENSGALG00000003030	DDX41	DDX41/ABSTRAKT	•	
ENSGALG00000003658	DHX35	DHX35	•	
hnRNP				
ENSGALG00000018665	-	hnRNP novel	•	
ENSG00000126457	HRMT1L2	HRMT1L2	•	
SR				
ENSGALG00000005525	SFRS1	SF2p33	•	
ENSGALG00000009484	SFRS5	SRp40	•	
ENSGALG00000002487	SFRS8	SFRS8	•	
RBP				
ENSGALG00000018992	RBM4B	RBM4B/Lark	•	
ENSGALG00000004626	RBM5	RBM5	•	
ENSGALG00000007017	RBM7	RBM7	•	
ENSGALG00000004555	RBM22	RBM22	•	
ENSGALG00000011038	CBX3	RNPS1	•	
ENSGALG00000006113	ZNF326	ZFP326	•	
ENSGALG00000010952	-	Requiem	•	
ENSG00000169564	PCBP1	PolyrCBP	•	
ENSGALG00000009427	TIAL1	TIA1	•	
ENSGALG00000001187	STRBP	STRBP	•	
Export/transcription/NMD				
ENSGALG00000004571	PPP1CA	GLC7/PPP1CA	•	
ENSGALG00000003220	RENT1	UPF1/RENT	•	
ENSG00000172660	TAF15	TAF15/RBP56	•	
3' end proc.				
ENSGALG00000010783	CPSF2	CPSF2	•	
ENSGALG00000004714	CPSF4	CPSF4	•	
ENSGALG00000013943	FIP1L1	FIP1	•	
Structural				
ENSGALG00000009126	TTN	Titin	•	
ENSG00000117245	KIF17	Kinesin KIF17	•	
Chromatin modification				
ENSGALG00000000360	ARID1A	ARID1A-SWI/SNF	•	
ENSGALG00000013683	ARID1B	ARID1B-SWI/SNF	•	
ENSGALG00000010164	SMARCA2	SMARCA2	•	
ENSG00000127616	SMARCA4	Brahma/SMARCA4	•	

Table 3.5 *(Continued)*

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]
ENSGALG00000009913	SMARCA5	SMARCA5	•	
ENSGALG00000005983	SMARCB1	SMARCB1	•	
ENSGALG00000005048	SMARCC2	BRG1-SWI/SNF	•	
ENSGALG00000005048	SMARCC1	SMARCC1	•	
ENSGALP00000010010	SMARCD1	SMARCD1	•	
ENSGALG00000000363	SMARCD2	SMARCD2	•	
ENSGALG00000002100	SMARCE1	SMARCE1	•	
Cyclophillins				
ENSGALG00000013383	PPIE	CYP-E	•	
ENSGALG00000004874	PPIH	USA-CYP	•	
ENSGALG00000014747	SDCCAG10	CYP16	•	
ENSG00000137168	PPIL1	PPIL1/CWF27	•	
ENSG00000100023	PPIL2	PPIL2/CYP60	•	
ENSG00000113593	PPWD1	PPWD1	•	

3.3.3 Compositional comparison of *in vivo* versus *in vitro* purified supraspliceosome

The core machinery, including snRNPs, SAPs, and SR proteins, is well represented in the material derived from all purification schemes. However, several polypeptides are exclusively contained in the supraspliceosome-associated material, or exclusively contained in the *in vitro* purified spliceosomes. We propose that this may result from the differences in procedure, or perhaps in the differences realized when analyzing a single intron-containing pre-mRNA versus all of the pre-mRNAs in a cell (Tables 3.1-3.4). We note that in the spliceosomes purified from *in vitro* extracts, many of the hnRNPs, RNA helicase-like proteins, and RNA binding proteins were identified. However, perhaps owing to specific binding of some polypeptides to the bulk pre-mRNA

and not the single transcript used in the *in vitro* spliceosome assembly reactions, a greater number of these polypeptides were present in the supraspliceosomes.

3.3.4 Pre-mRNA processing factors not present in any spliceosome purification

To estimate the aggregate number of polypeptides participating in the nuclear pre-mRNA processing pathway between all of the purifications, we have compiled a list of factors know to function in post-RNA Pol II transcript processing, but not present in any of the five purifications listed in Tables 3.1-3.5. In Table 3.6, we outline this small list of factors. The inability to be detected may be due to transient interaction with spliceosome, low abundance or correlated with the presence of an abundance hnRNP protein of similar molecular weight. We have not included factors for which there are yeast homologues, but no identifiable human or vertebrate homologue in the genomic databases.

Table 3.6 Known pre-mRNA processing proteins not present in any purified splicing complex.

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]
ENSG00000095485	CWF19L1	CWF19
ENSG00000152404	CWF19L2	CWF19
ENSG00000165630	PRPF18	PRP18
ENSG00000140829	DHX38	PRP16
ENSG00000149532	-	CFI-59K
ENSG00000165494	PCF11	PCF11
ENSG00000172409	CLP1	CLP1
ENSG00000111880	HCAP1	HCE/CEG1
ENSG00000146007	ZMAT2	SNU23
ENSG00000108296	CCDC49	CWC25
ENSG00000101138	CSTF1	CSTF-50
ENSG00000101811	CSTF2	CSTF-64
ENSG00000161981	C16ORF33	U11/U12-25K
ENSG00000184209	-	U11/U12-35K

3.3.5 Validation of ZFR as a *bona fide* spliceosome component

To demonstrate the authenticity of a novel polypeptide that co-purified with endogenous spliceosomes, we generated antiserum against ZFR (Table 3.3) and used it to specifically immunopurify ZFR-associated components. In Figure 3.6A, we show that the ZFR polypeptide is present in very high molecular weight complexes that co-migrate with supraspliceosomal material. In Figure 3.6B, we show that the anti-ZFR antiserum, but not the pre-immune serum or the Protein-A beads, immunoprecipitates the U1, U2, U4, U5 and U6 snRNAs. As a positive control, we showed that antiserum directed against the known spliceosomal protein SR140, prepared and analyzed under identical conditions, also immunoprecipitated all of the snRNAs. We also tested for the presence of another pre-mRNA splicing factor, hPrp43 (DHX15), in the material immunopurified with anti-ZFR; Figure 3.6C shows that the specific antiserum, but not the pre-immune serum or the Protein-A beads, immunoprecipitates hPrp43p/DDX15. This demonstrates that the novel spliceosome-associated factor ZFR is indeed associated with spliceosomal snRNAs and other spliceosomal proteins.

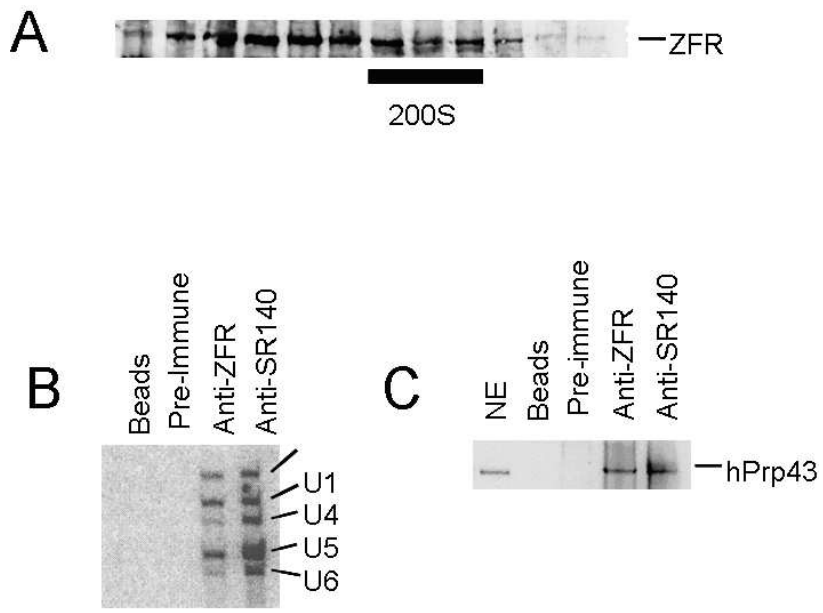

Figure 3.6 The novel Zn finger protein ZFR is a *bona fide* spliceosomal component.

(A) ZFR sediments with the 200S particle in glycerol gradients. HeLa nuclear extract was subjected to glycerol velocity gradient sedimentation analysis as in Figure 2. Proteins from the indicated fractions were electrophoresed through SDS-PAGE gels and subjected to western blot analysis using anti-ZFR antiserum. The bar below the gel denotes the 200S region. (B) ZFR is specifically associated with spliceosomal snRNAs. Equal amounts of HeLa nuclear extract were incubated with protein-A beads (beads), pre-immune serum and protein-A beads (pre-immune), anti-ZFR antiserum and protein-A beads (anti-ZFR) or anti SR140 antiserum and protein-A beads (anti-SR140) according to the Materials and Methods. Recovered nucleic acids were subjected to northern blot analysis and probed with antisense probes to human snRNAs (identities noted to the right of the Figure). (C) ZFR is specifically associated with complexes containing spliceosomal proteins. Immunoprecipitation conditions and lanes are as described in (B). Proteins were subjected to western blot analysis using hPrp43 antiserum.

3.4 DISCUSSION

Affinity purification of the SmD3-associated proteins from DT40 cells has allowed us to probe more deeply into the general pre-mRNA processing machinery present in vertebrate cells. Indeed, when combined with the data from *in vitro* assembled spliceosome characterization and known splicing factors not detected in any complexes previously purified, we show there are at least 305 polypeptides involved in or present during the processing of nuclear pre-mRNA. In Figure 3.7, I present a schematic model of the pre-mRNA processing pathway *in vivo* that encompasses the concept of the supraspliceosome.

What is perhaps most remarkable about our results is the fact that, despite the operationally distinct purification strategies, the basal pre-mRNA processing machinery required to effect the removal of a single intron *in vitro* is not significantly different than that purified from complex mixtures of all of the pre-mRNAs in a vertebrate or human nucleus. The major differences in composition between the previous purifications and the one described herein involve 1) polypeptides predicted by sequence homology to interact with the pre-mRNA 2) the depth of coverage for polypeptides involved in export and 3' end processing, and 3) polypeptides which may require that the pre-mRNA in these complexes follow the path of RNA Pol II transcription and nuclear trafficking, such as the SWI/SNF complexes, structural proteins and nucleoporins. Two possible classes of polypeptides may exist that are not detected in our preparations. First are those that are underrepresented because they may interact with only a small number of pre-mRNAs, such as intron-or exon-specific binding proteins. The other class of proteins which may participate in pre-mRNA splicing but is absent from our analysis might include tissue-specific developmental stage-specific factors which would not be present in our bulk supraspliceosome preparation due to the use of only two cells types.

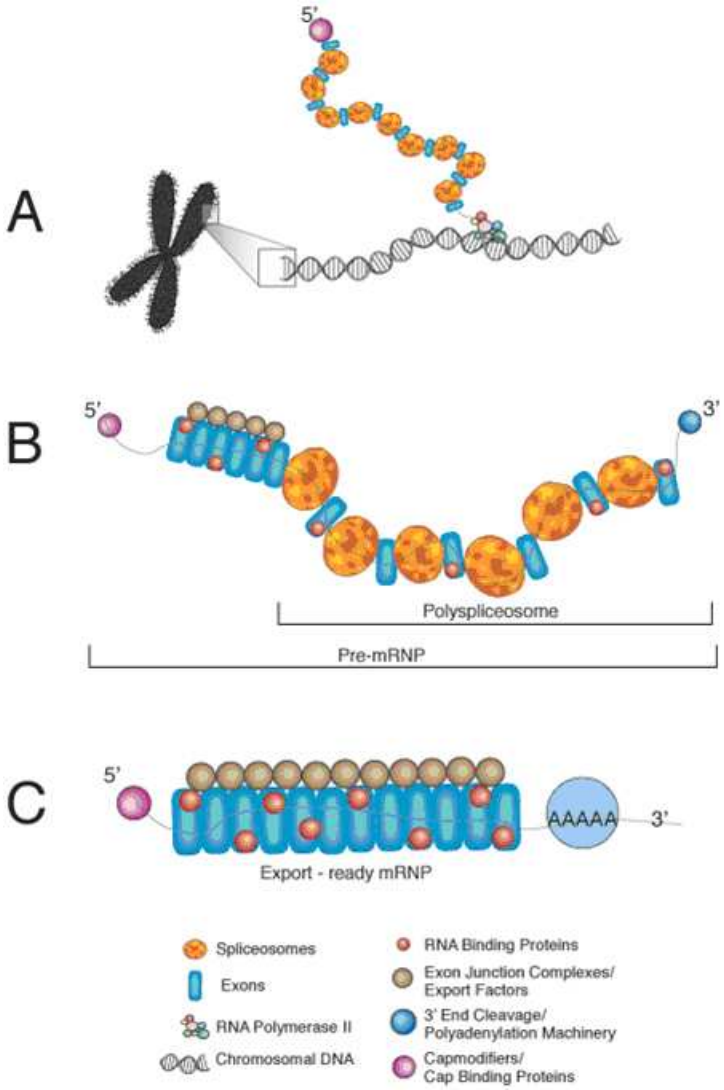

Figure 3.7 Model of the vertebrate supraspliceosomes in gene expression.

(A) Co-transcriptional assembly of spliceosomes, 5' end modification machinery and other pre-mRNA binding factors with RNA polymerase II transcripts. (B) Released transcript contains numerous spliceosome moieties, spliced and un-spliced exons and 5' cap-binding complex and 3' end processing factors. (C) Mature mRNA contains RNA binding proteins, 5'- and 3'-end stabilizing factors, and proteins which promote the export of the mRNA.

Our mass spectrometry peptide data revealed several intriguing and novel polypeptides in these complexes. We found chicken homologues of the yeast splicing factors Prp38p, Prp39p and Aar2p, previously unannotated in purified pre-mRNA splicing complexes.

A number of uncharacterized RNA binding proteins were identified in our purifications. Recent studies have revealed the functions of some of the RNA binding protein. RBM4 has been shown as alternative splicing regulator that favors intron inclusion (Kar et al., 2006; Lai et al., 2003). Guo *et al*. demonstrated that RBM7 interacted with splicing factor 3b subunit 2 (SAP145), and with the splicing regulator, SRp20a, suggesting a role in pre-mRNA processing pathway (Guo et al., 2003).

One of the novel polypeptides of interest in the endogenous splicing complexes is RAT1p/XRN2. Rat1p has been identified as a nuclear 5' to 3' exonuclease (Amberg et al., 1992; Johnson, 1997; Kim et al., 2004) and involved in many RNA processing pathways such as rRNA processing, snoRNA biogenesis, mRNA export, and transcription termination. Interestingly, when our lab characterized the temperature sensitive mutant *prp27-1*, initially identified from Abelson lab (Vijayraghavan et al., 1989), it was shown to be allelic to *RAT1* (Gupton et al., submitted). This *prp27-1* mutant resulted from a single amino acid change in a conserved yet uncharacterized domain. At non-permissive temperature, this mutant causes severe post-splicing intron-accumulation phenotype, suggesting its role in spliceosome disassembly or intron turnover pathways. A bank of *rat1* mutants has been created to further determine the domain specificity to intron turnover. Another novel polypeptide, Tra2 collaborates with other SR proteins and regulate alternative splicing of gonadotropin releasing hormone pre-mRNA (Paradis et al., 2007; Park et al., 2006).

Furthermore, two uncharacterized AAA ATPases, ATAD3A and ATAD3B that may perform chaperone-like functions and assist assembly or disassembly of protein complexes were identified in both chicken and human supraspliceosome. We report the identity of all the novel factors in Table 3.7

Although we cannot completely eliminate the possibilities that there may be contaminants present in our preparations, our results are validated by evidence from our lab and other researchers that proteins Brahma, Rat1, and Tra2α have a novel role in splicing that were not detected previously in the *in vitro*-assembled spliceosome.

With the introduction of the CLEP tagging technology to novel proteins in DT 40 cells or to more cell types (Chen et al., 2006) and using small interfering RNA (siRNA) for gene silencing (Paul et al., 2002) in other cell lines or creating knockout in DT40 cells (Winding and Berchtold, 2001), it will be possible to characterize these factors more rapidly.

Table 3.7 Novel polypeptides present in the supraspliceosomes and *in vitro* assembled splicing complexes

ENSEMBL accession #[a]	HGNC[b]	Polypeptide[c]	Gg PS[d]	Hs PS[e]	N[f]	R[g]	Z[h]
Novel or unknown to splicing							
ENSGALG00000011351	HSP90AA1	HSP90α	•	•			
ENSGALG00000010175	HSP90AA2	HSP90ß	•	•			
ENSGALG00000012726	HSP90B1	HSP108	•	•			
ENSGALG00000009967	LRPPRC	LRP130	•				
ENSGALG00000003693	MACF1	Macrophin	•				
ENSGALG00000007705	NCL	Nucleolin	•				
ENSGALG00000015933	C21ORF66	GCRBF	•				
ENSGALG00000008454	-	NOP58	•				
ENSGALG00000009061	ACTL6A	BAF53A	•				
ENSGALG00000007520	SSRP1	FACT80	•	•			
ENSGALG00000015821	CCT8	TCP1-theta	•				
ENSGALG00000014500	NOL5A	Nol5A/NOP56	•				
ENSGALG00000005624	MED12	TRAP230	•				
ENSGALG00000010973	TRA2A	TRA2a	•	•			
ENSGALG00000008372	XRN2	RAT1	•	•			
ENSG00000197157	-	SND1	•	•			
ENSGALG00000001948	SAFB2	SAFB/HSP27	•	•			
ENSGALG00000003177	BRD8	BRD8	•	•			
ENSGALG00000010699	-	FOG	•				
ENSGALG00000005177	C9ORF10	C9ORF10	•	•			
ENSGALG00000002653	-	ELG	•	•			
ENSGALG00000016949	WBP4	WBP4	•				
ENSGALG00000004133	FUSIP1	Fus IP	•	•			
ENSGALG00000017384	ERH	ERH	•	•			
ENSG00000079246	XRCC5	Ku80		•			
ENSG00000182562	ATAD3A	ATAD3A (AAA ATPase)	•	•			
ENSGALG00000001515	ATAD3B	ATAD3B (AAA ATPase)	•	•			
ENSG00000108588	CCDC47	CCDC47		•			

Chapter 4: Gene targeting in mouse embryonic stem cells and generation of mice carrying an epitope tag

4.1 BACKGROUND

Mouse embryonic stem cells are pluripotent cells derived from the inner mass of preimplantation blastocyst-stage embryos. Bradley *et al*. showed that ES cells had the capacity to colonize the reproductive organs of chimeric mice and transmit genetic information to progeny even after having been cultured for a long period of time *in vitro* (Bradley et al., 1984; Robertson et al., 1986). This observation paved the way for the generation of transgenic mice by manipulating genes in ES cells. Alteration of the mouse genome by homologous recombination in ES cells was first achieved in 1987. Doetschman *et al*. and Thomas *et al*. were able to show that by homologous recombination, segments of the mouse genome could be disrupted or replaced (Doetschman et al., 1987; Thomas and Capecchi, 1987). The ability of ES cells to enter the germline after genetic manipulation in culture and differentiate into different cells types depends on the conditions in cell culture. To keep the ES cells in an undifferentiated state in the cell culture system, they are co-cultured with a feeder layer of cells, typically mouse embryonic fibroblasts coated on the inner layer of culture dishes and/or the addition of leukemia inhibitory factor to the culture media (Hogan, 1994; Pease and Williams, 1990). The feeder layer not only provides cellular contact for the maintenance of the ES cells in the undifferentiated state but also releases nutrients to the media. Transgenic mice, knockout and knockin mice, generation is based on ES cell-derived techniques. Coupled with Cre/lox-based approaches, tissue-specific and conditional knockouts can be obtained, which are particularly useful with embryonically lethal genes. Gene-targeting has become a feasible tool to produce mice with alterations

in specific genomic loci. It is an important tool to study the function of individual proteins, biological complexes and pathways [see review in (Muller, 1999)].

In this chapter, I show how I applied the CLEP technique (described in Chapter 2) to mouse embryonic stem cells and describe the transgenic mice that were generated. I created two targeting constructs for two different spliceosomal proteins for gene targeting in mouse ES cells. First, I inserted a TAP tag at the carboxyl-terminus of the U4/U6•U5 (tri-snRNP) associated protein, SART-1 (squamous cell carcinoma antigen recognized by T-cells 1; also known as U4/U6•U5-110K). Secondly, I chose a U5-associated protein U5-220K and inserted a polyhistidine tag at its carboxyl-terminus.

SART1 was identified previously as a gene encoding a protein recognized by cytotoxic T cells of squamous cell carcinoma patients (Shichijo et al., 1998). Expression of SART1 might be a potential tool for specific immunotherapy of patients with tumors. It is also identified as an autoantigen which can interact with IgE autoantibodies from patients suffering from actopic dermatitis (Valenta et al., 1998). Subsequently, Makarova *et al.* showed that SART1 was an SR-related U4/U6•U5 tri-snRNP protein. SART-1 encodes a protein of 800 amino acids and a predicted molecular weight of 90,263 kDa; however, the slower migration behavior of SART1 in SDS-PAGE gels, corresponding to an molecular weight of 110 kDa was assumed due to the glutamic acid rich region comprising amino acids 450-600 (Makarova et al., 2001). It is orthologous to the yeast Snu66p and is important for the recruitment of the U4/U6•U5 tri-snRNP to the pre-spliceosome and the assembly of mature spliceosome (Stevens and Abelson, 1999)

U5-220K, an ortholog of yeast Prp8p, is the largest and most highly conserved U5 snRNP associated protein which in the course of splicing contacts U5 snRNA, U6 snRNA, the branchpoint of pre-mRNA, the 5’ and 3’ splice sites in the intron as detected

by photochemical cross-linking. It interacts with several splicing proteins and is involved in both the first and second step of splicing [reviewed in (Grainger and Beggs, 2005)].

The mice carrying CLEP tags will not only allow us to characterize spliceosomes from different tissues but also perform biochemical and histological analyses from the native, endogenous pre-mRNA splicing machinery in a living animal.

4.2 MATERIALS AND METHODS

4.2.1 Targeting vector construction

The targeting vector creation is described as in section 2.2.1. The oligonucleotides for creating the SART1-TAP targeting vector and U5-220K-$(His)_8$ are shown in Table 4.1.

Table 4.1 Oligonucleotides used for targeting vector construction in ES cells

Targeted gene	Oligo sequence
mSART-1A	5'- GCGGCCGCGAAATAGGAGCAGCTGTGAACATGG-3'
mSART-1B	5'- TCTAGAGGATCCTTTGGTGATGGTGTTCCTGCAGGG-3'
mSART1C	5'-GTCGACAGCCGCCCTCCTCCCTGGCCCAGATG-3'
mSART1D	5'-TTAATTAAGATCCACAGGCACAGCCGAGAACAC-3'
mSART1E	5'-ATCGATTGGCTAGTGATGGTTGTGCAGGAGTG-3'
mSART1F	5'-TGATCAATTGACCCATTCTGTCAATGGGTGTG-3'
mU5-220KA	5'-GCGGCCGCGGCTAAAAATATTTTTGTATGTAGG-3'
mU5-220KB	5'-AAGCTTCTAGTGGTGATGATGATGATGATGATG GGCATAGAGGTCCTCTCTGTCTGCAG-3'
mU5-220KC	5'-GTCGACTTACTTCCTTCTGCTTCAGATCCCC-3'
mU5-220KD	5'-TTAATTAACGGGTCCGAAGCGGCGTGCCAGCCC-3'
mU5-220KE	5'-CTCGAGACGCGGCGGCTGGGCTGGTGCCACTAG-3'
mU5-220KF	5'-GCTAGCTCCAGTCCCTCACCGCTGGAAGTACGC-3'

4.2.2 ES cell electroporation and generating transgenic mice

To generate transgenic mice using mouse ES cells, SM1 mouse embryonic stem cells (129S6) were electroporated with the linearized targeting construct. ES clones which survived in double selection media containing G418 and ganciclovir were isolated, expanded into 6-well plates and grown until confluent. PCR analysis of purified genomic DNA (DNeasy tissue kit, Qiagen) was performed to screen for epitope-tag insertion into the correct chromosomal locus. To generate transgenic mice (Illustration 4.1), correctly targeted ES clones were microinjected into E3.5 C57Bl/6 blastocysts and implanted into (CD1) pseudopregnant females. Resulting chimeric males were mated to C57Bl/6 females for germline transmission of the altered allele. The resulting agouti mice were screened for the SART1-TAP with western blot analysis. Male and female heterozygotes carrying the epitope tag in one of the alleles are mated to one another to generate homozygotes carrying TAP tags in both alleles.

All mice were housed in accordance with protocols approved by the Animal Care and Use Committee of the University of Texas at Austin.

4.2.3 Screening for the epitope-tagged gene by PCR and RT-PCR analysis

Genomic DNA was extracted (DNeasy tissue kit, Qiagen) from G418- and ganciclovir-resistant ES colonies to screen mouse ES cells positive for mSART-1-TAP. PCR reactions were performed using one oligonucleotide designed outside the targeting construct region (PCR5) and the other one inside the targeting vector (PCR3): mSARTPCR5 (5'-CCTCTCACTGCCTCTCCTGCTGCGGGG-3') and mSARTPCR3 (5'- GAGGAGGCTCAGCGACTTGCCAGATGG-3'). The genotypes of agouti mice and progeny were screened for evidence of targeted insertion of the TAP tag using RT-PCR analysis from whole blood total RNA (Mouse RibopureTM-Blood RNA Isolation

Kit, Ambion) with oligonucleotides (RTPCR5) mSARTPCR5 (5'-GCTCCAGGAGAAGCAGAAGGC-3') and (RTPCR3) mSARTRTPCR3 (5' CTAACCTTAGACAGCAGCACAACCAAGAG3'). The oligonucleotides for screening the insertion of polyhistidine tag into U5-220K gene were mU5220KRTPCR5 (5'-CAACAACCCCAAAGGCTACCTACCCTCGC-3') and mU5220KRTPCR3 (5'-GTTCAAGAGGCTCCAGGCTTTGGGGATC-3').

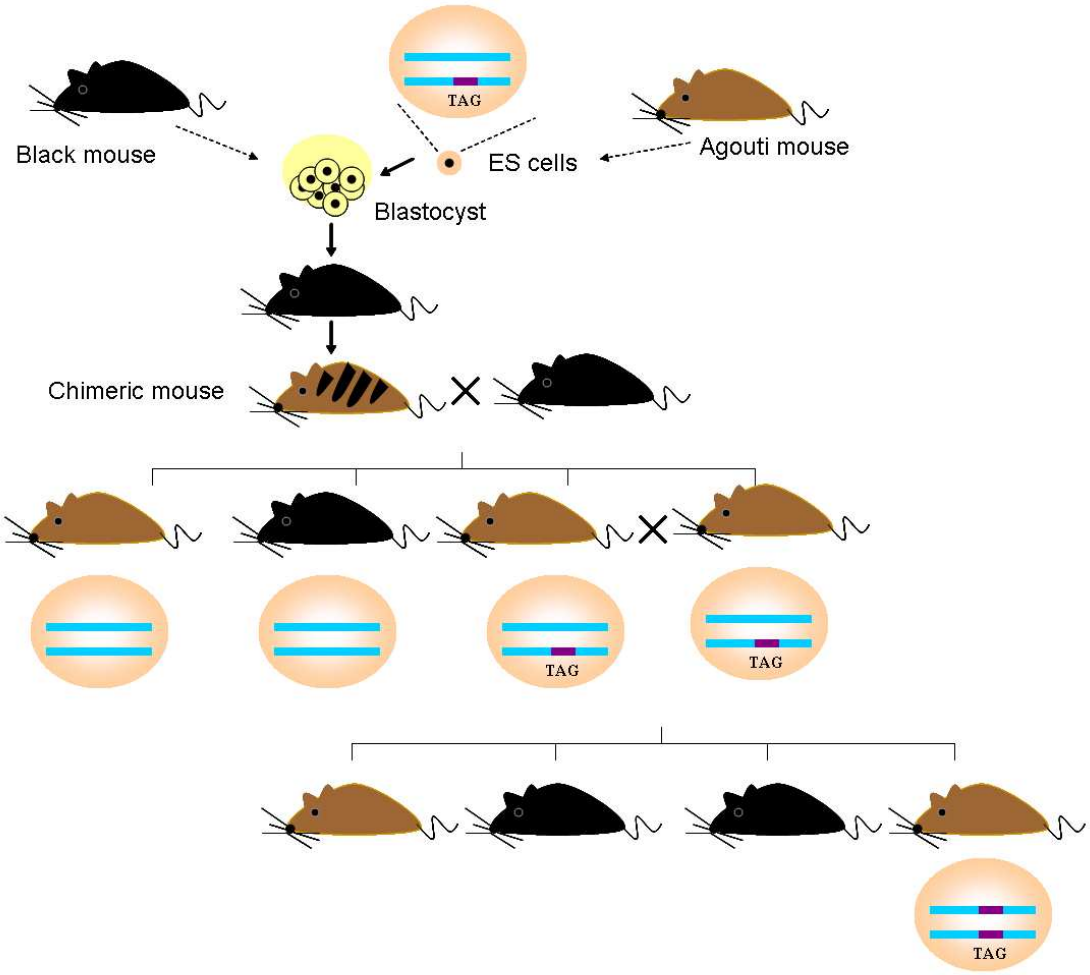

Illustration 4.1 Producing transgenic mice carrying an epitope tag

Embryonic stem cells carrying a targeted epitope tag are injected into the blastocyst-stage embryo from a black mouse strain and then placed in surrogate mothers. If the ES cells can survive and proliferate in the embryos, the newborns will become chimeric carrying cells from two mouse strains. Chimeric mice are then mated to black mice and the resulting agouti progeny are screened for evidence of the targeted insertion of the epitope tag. Male and female heterozygotes carrying an epitope tag in one of the alleles are mated to one another to generate homozygotes carrying epitope tags in both alleles.

4.2.4 Western blotting analysis

Mouse ES cells and organs from sacrificed mice were processed by homogenization in LDS sample buffer. Approximately 50 mg of each organ was homogenized in 1 ml of SDS-PAGE loading buffer. The agouti mice expressing SART1-TAP were confirmed by western blot analysis from tails homogenized in the LDS sample buffer. Homogenized samples (20 µl) were electrophoresed through SDS-PAGE gels and transferred to a nylon membrane for western blotting analysis. Peroxidase-antiperoxidase (PAP) antibody (Sigma) was used for detecting TAP tagged proteins.

4.2.5 Preparation of extracts from liver nuclei and brain nuclei

The preparation of extracts was performed as previously described (Lu et al., 1998) with some modifications. Four livers and four cerebra were collected on ice, washed with 1X PBS and cut into small pieces in ice-cold buffer A [1.3 M sucrose, 1 mM $MgCl_2$, 10 mM potassium phosphate, (pH 6.8)]. The small pieces were homogenized into 8 vol of buffer A containing 0.2 mM PMSF, 1 µg/ml leupeptin, and 1 µg/ml pepstatin in a dounce homogenizer with 10 strokes on ice. The homogenate was filtered through 2 layers of cheesecloth and centrifuged at 1000 x g for 20 minutes. The resulting pellets were then suspended in a 2 ml of ice-cold buffer A containing 0.2 mM PMSF, 1 µg/ml leupeptin, and 1 µg/ml pepstatin and homogenized again in a dounce homogenizer with 10 strokes. The resulting suspension was mixed with 3.5 vol of ice-cold buffer B [2.4 M sucrose, 1 mM $MgCl_2$, 10 mM potassium phosphate, (pH 6.8)] and centrifuged at 100,000 x g for 1 hour. The white nuclear pellet was suspended in 2 ml of low-salt buffer (30 mM Tris-HCl, 125 mM KCl, 5 mM $MgCl_2$, 0.5% Triton-X100, 0.15mM spermine, 0.05mM spermidine, and 0.2 mM PMSF), and sonicated at the

maximum output, twice for 20 seconds on ice with 1 minute in ice between sonications. The sonicated mixture was centrifuged at 5000 x g for 15 seconds and the supernatant, the nuclear fraction, was used for TAP purification as described in section 2.2.7.

4.3 RESULTS

4.3.1 Live mammals carrying CLEP tags

4.3.1.1 Mice carrying SART1-TAP

After construction and clone validation, the targeting vector is linearized at an appropriate site upstream of the AB fragment (usually NotI) or downstream of the EF fragment (if not using ganciclovir selection). The SART1-TAP targeting vector was linearized with NotI and electroporated into mouse ES cells. After selection with G418 and ganciclovir, one hundred forty one resistant colonies were selected and expanded. The presence of correctly targeted cells was tested by PCR using one oligonucleotide primer designed outside the targeting region (PCR5 in Figure 2.1) and one oligonucleotide primer designed inside the targeting vector (PCR3 in Figure 2.1). Seven out of 141 ES cell clones were positive for the SART1-TAP construct (Figure 4.1A) and all seven were positive for the TAP tag of an appropriately sized band (~130 Kd) by western blotting (Figure 4.1B). Three chimeric mice were produced from injected blastocysts and tissue samples from brain, heart, kidney and spleen were analyzed for SART1-TAP by western blotting. While wild-type mice were negative for TAP signal, the chimeric mice were positive in all tissues analyzed (Figure 4.1C).

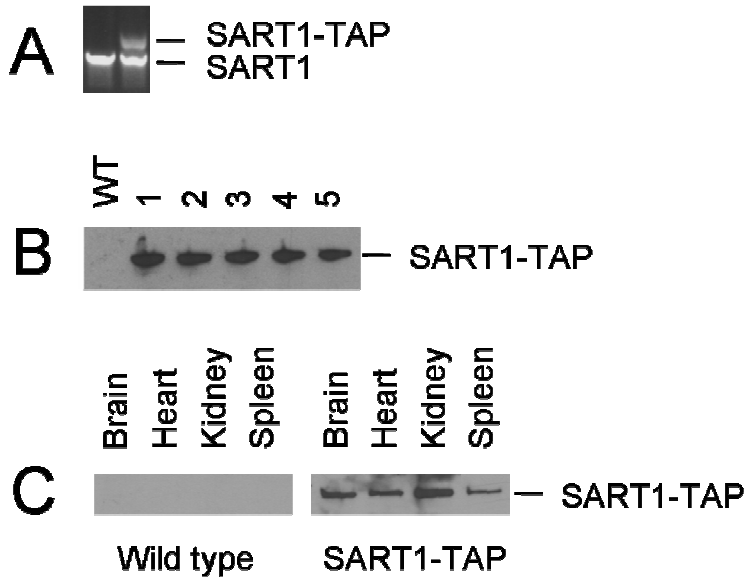

Figure 4.1 Targeted SART1-TAP in ES cells and live mice.

(A) Detection of SART1-TAP targeted mouse ES cell clones by PCR. Wild-type and targeted alleles are noted. (B) Western blot analysis of 5 of the 7 mouse ES cell clones positive for SART1-TAP from [A]. The 130Kd SART1-TAP is noted. (C) Western blot analysis of SART1-TAP from organs of the mice generated from the mouse ES cells. Left panel, wild-type untagged mouse organs, right panel, SART1-TAP mouse organs.

Chimeric mice were mated to black mice which resulted in 11 agouti progeny. Six out of eleven agouti mice showed positive signals for SART1-TAP of an appropriately sized 130KDa band (Figure 4.2A). The genotype of progeny from agouti mice was determined by RT-PCR from whole blood RNA using one oligonucleotide primer designed outside the targeting region and one inside the 3'UTR which is the same in the targeted and wild-type loci. The wild-type mice contain two wild-type alleles, the heterozygotes contain one wild-type allele and one targeted allele, which is 590-nucleotides larger due to the tag and the homozygotes contain two targeted alleles (Figure 4.2B).

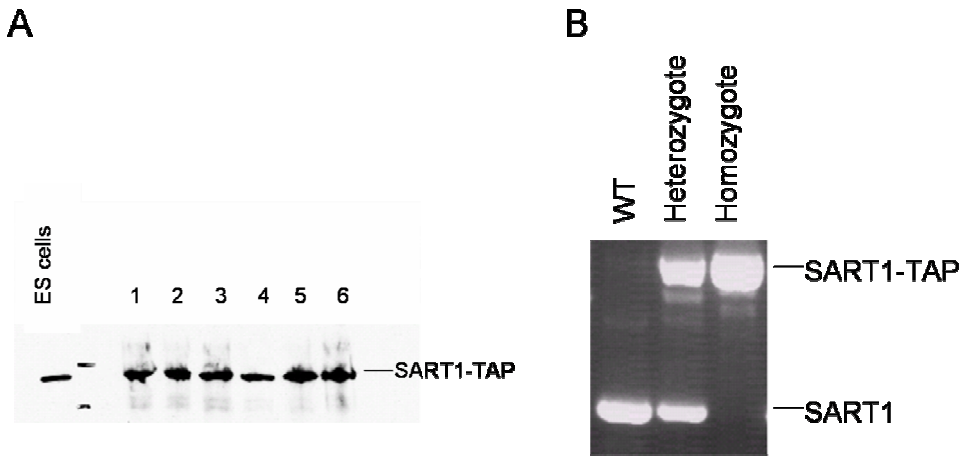

Figure 4.2 Targeted SART1-TAP in agouti mice and their progeny.

(A) The genotypes of agouti mice were determined by western blot analysis from tails homogenized in LDS sample using peroxidase-antiperoxidase (PAP). We generated 11 agouti mice; six out of eleven agouti mice showed positive signals of SART1-TAP.
(B) RT-PCR analysis of progeny from SART1-TAP transgenic mice.
Oligonucleotide primers designed upstream of the stop codon and within the 3'UTR were used to identify wild-type, heterozygote SART1-TAP or homozygote SART1-TAP mice.

4.3.1.2 Mice carrying U5-220K-$(His)_8$

Utilizing the method described above, a polyhistidine tag $(His)_8$ was inserted into the carboxyl terminus of U5-220K. Two out of one hundred and thirty one ES cell colonies were positive for a HIS tagged U5-220K by RT-PCR of RNA (Figure 4.3A) and 1 of 3 *agouti* progeny from HIS-tagged chimeras showed insertion of HIS-tag in one of the chromosomes (Figure 4.3B). Unfortunately, by sequencing analysis, insertion of the polyhistidine tag inhibited the splicing of the last intron of U5-220K, leading to the accumulation of U5-220K mRNA containing the last intron sequence. U5-220K mRNA

containing the last intron sequence was not able to express polyhistidine tag for purification due to the presence of a stop codon within the intron sequence.

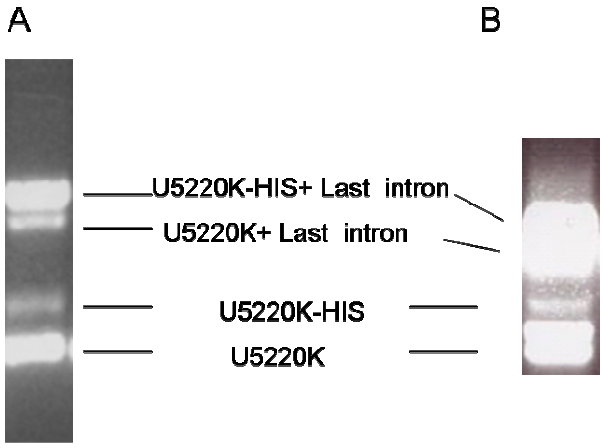

Figure 4.3 Targeted U5-220K-$(His)_8$ in ES cells and in agouti mice.

Correctly targeted U5-220K-$(His)_8$ were screened by RT-PCR analysis. The PCR fragments containing the last intron sequence were noted.

4.3.2 Purification of SART1-associated proteins from brain and liver cells

Splicing regulation, particularly alternative splicing, in a specific tissue or developmental stage contribute to functional complexity in the genome (Xu et al., 2002). Epitope tagging of proteins at the native chromosomal loci of genes in mice provides a useful tool to study the expression and organization of the gene expression machinery during different stages of development and in different tissues. In addition, nuclear extracts derived from differentiated tissue serve as a useful tool to identify regulatory processing factors that are physiologically significant. By purifying pre-mRNA processing complex from the livers of SART-1-TAP tagged homozygote, we will produce the *in vivo* assembled particles, which will provide great information and application to the study of tissue-specific regulatory mechanisms.

Liver and brain cells were first used for purification because of their relatively large biomass. Moreover, other researchers have successfully purified proteins from liver cells or prepared splicing-active extracts from brain cells in mice/rats (Grabowski, 2005; Lu et al., 1998). Using similar nuclei extraction methods combined with TAP purification, I tried to purify SART1-associated proteins from brain and liver. For brain samples, after running purified fractions on the SDS-PAGE gels, the protein bands looked exactly the same as that from wild type, non-tagged mice, suggesting no polypeptides specifically associated with SART1 were purified with this purification method (Figure 4.4); however, according to my western blotting analysis of brain tissue samples, I was able to identify the presence of SART1-TAP proteins in brain (Figure 4.1C). I conclude that SART1 proteins are not abundant enough for purification in brain tissue but can be detected by western blot analysis. This result is supported by the database of gene expression pattern in mouse brains from Allen Institute for Brain Science/Allen Brain Atlas, as no SART1 expression was noted by in situ RNA hybridization (Figure 4.5). Alternatively, the yeast homolog of SART1, Snu66p, is not essential in yeast but disruption of *SNU66* confers a cold-sensitive phenotype (Stevens et al., 2001). This data suggests that SART1 may not be an essential splicing factor in brain. For the liver samples, I was able to purify the polypeptides specifically interacting with SART1 and U4, U5, and U6 snRNAs were present in the purified components (Figure 4.6). These polypeptides will further be confirmed and validated by mass spectrometry.

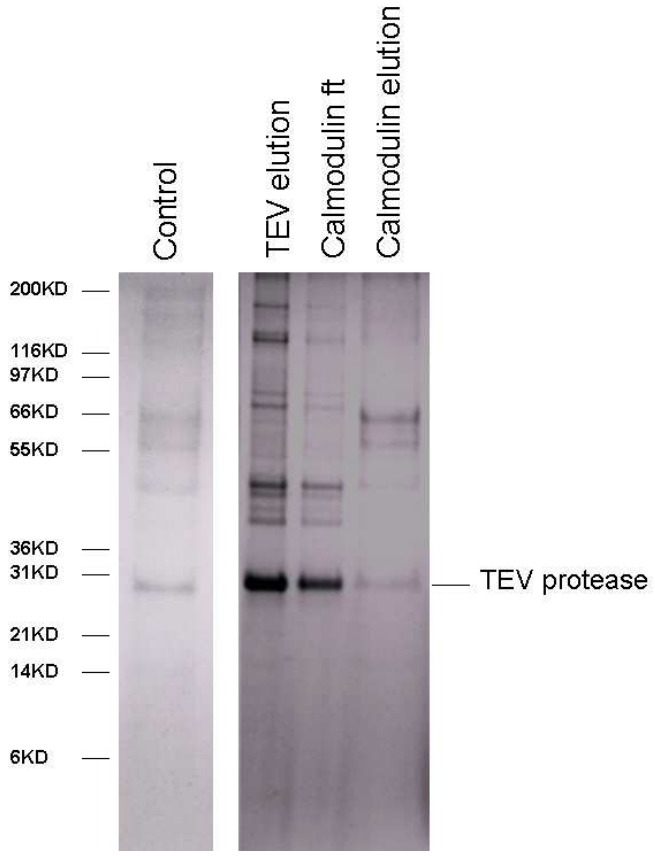

Figure 4.4 SART1-associated polypeptides in mouse brain.

Calmodulin elution from a mock purification in which WT type (non-tagged) mice brain samples was subjected to the TAP purification as background control in lane 1. TEV elution, calmodulin beads flow through, and calmodulin elution from SART1-TAP homozygote brain samples were loaded in lane 2, 3, and 4, respectively.

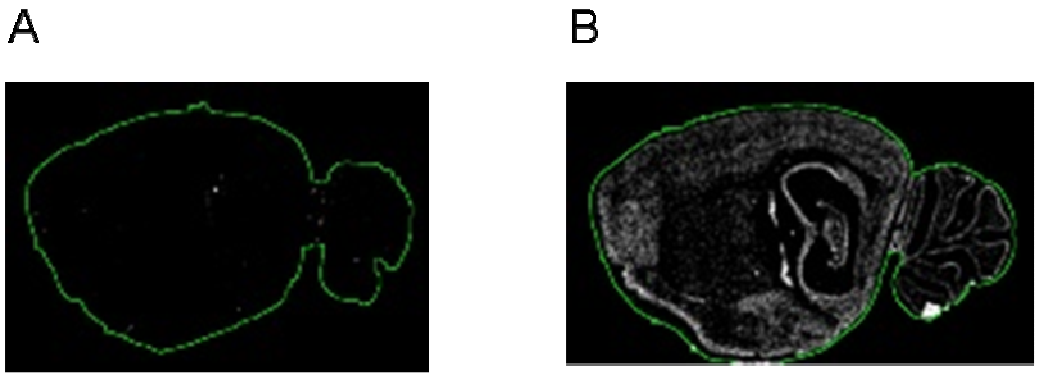

Figure 4.5 SART1 and Prp8f expression in brain.

SART1 (A) and Prpf8 (B) expression level were detected by RNA in situ hybridization. Expression of SART1 was not detected compared to that of another splicing factor Prpf8. The images were adapted from:
http://www.brain-map.org:80/viewImage.do?imageId=76074004&coordSystem=pixel&x=5088.5&y=3464.5&z=7.86086272968458&initExp=y
http://www.brain-map.org:80/viewImage.do?imageId=68250086&coordSystem=pixel&x=5384.5&y=3240.5&z=7.42873061565605&initExp=y

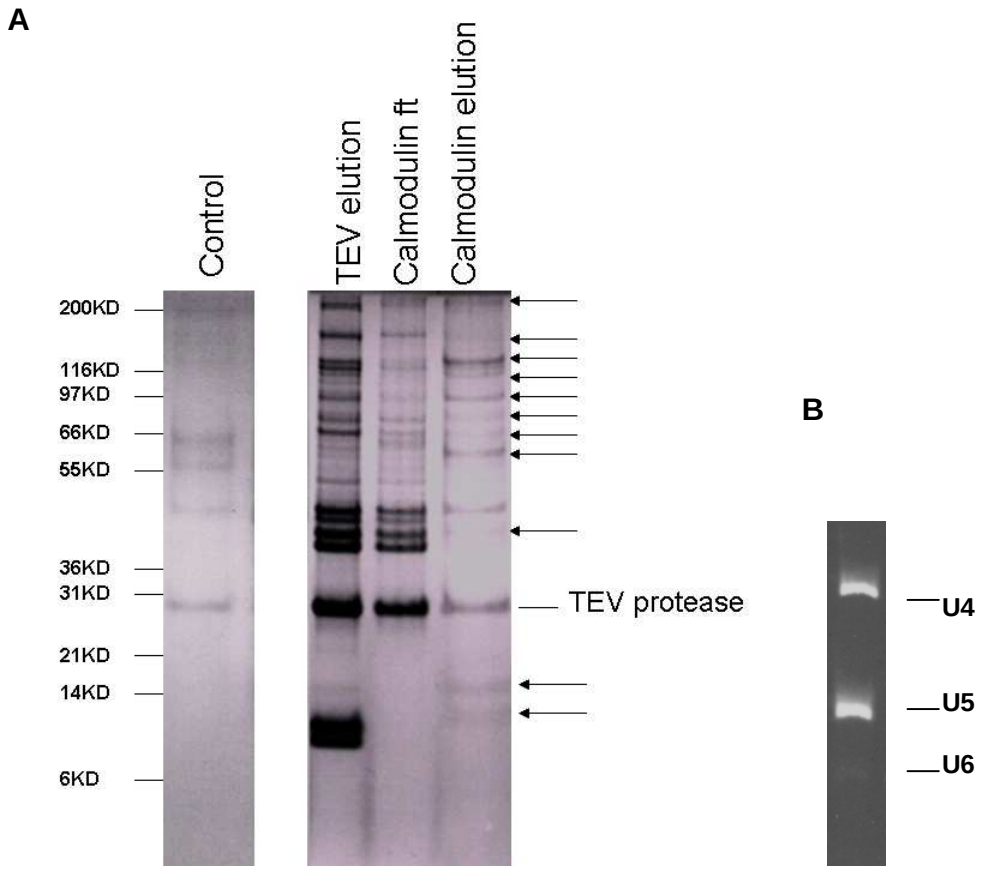

Figure 4.6 SART1-associated polypeptides and RNA in mouse liver.

(A) Calmodulin elution from a mock purification in which WT type (non-tagged) mice liver samples were subjected to the TAP purification as background control in lane 1. TEV elution, calmodulin beads flow through, and calmodulin elution from SART1-TAP homozygote liver samples were loaded in lane 2, 3, and 4, respectively. Polypeptides present only in the calmodulin fractions but not in the control were noted as arrows. (B) U4, U5, and U6 snRNAs were noted in the purified complex.

4.4 DISCUSSION

These CLEP procedures will be of great use where circumstances require that the gene of interest be controlled by its native promoter, especially in live animals where expression of proteins at inappropriate levels can lead to developmental defects or unwanted interactions. CLEP tagged animals should also prove useful for the study of virtually any process and serve as a biochemical tool to study the organization of protein complexes during different stages of development and in different tissues, especially those of disease relevant gene products.

Chapter 5: Induction of a Severe Autoimmune Response by Modification of SART1 (U4/U6•U5-110K) in Mice

5.1 BACKGROUND

Inappropriate autoimmune responses result from the breakdown of self-tolerance in which the immune system is not able to distinguish self antigens from foreign antigens. In healthy individuals, several tolerance mechanisms, working independently but in concert, maintain the balance between activation of inflammatory responses and prevention of autoimmunity (Gregersen and Behrens, 2006). These mechanisms include clonal anergy, clonal deletion and peripheral suppression, such as immune suppression by natural T regulatory (Treg) cells and dysregulation of cytokines (Heath et al., 1996; O'Shea et al., 2002; Takahashi and Sakaguchi, 2003).

Over the past few years, the discovery of functionally distinct subsets of lymphocytes has expanded our knowledge regarding the maintenance of self-tolerance. For example Treg cells, expressing the cell surface markers CD4 and CD25 and the transcription factor FoxP3, play a critical role in suppression of the activity of effector T cells (Asano et al., 1996; Fontenot et al., 2003; Randolph and Fathman, 2006). Imbalances in effector and regulatory T cells leads to autoimmune diseases (Torgerson, 2006). Another subset of lymphocytes, the marginal-zone (MZ) B cells, is located at the blood-lymphoid interface. These cells are enriched with self-reactive clones and are proposed to initiate autoantibody production (Wellmann et al., 2001; Zeng et al., 2000).

Increases in MZ B cells occur in lupus-prone mice even before clinical manifestation of the disease (Wither et al., 2000). Multiple activated pathways contribute to the process of autoimmune disease and inappropriate immune response is often the cause of organ-specific or systemic damage (Marrack et al., 2001). The cause

of autoimmune diseases is not fully understood, but in some cases it can be triggered by exposure to drugs, microorganisms, or by environmental causes in people with a genetic predisposition to autoimmunity (Marrack et al., 2001).

A high titer of autoantibodies is a diagnostic hallmark in many autoimmune diseases (von Muhlen and Tan, 1995). Anti-nuclear antibodies, including several populations of antibodies with different specificities, have been identified in patients with various autoimmune diseases (von Muhlen and Tan, 1995). Among these anti-nuclear antibodies, is a spectrum of antibodies to small nuclear ribonucleoproteins (snRNP) involved in pre-mRNA splicing, such as anti-Sm and anti-U1 RNP (Hoet et al., 1993). Non-spliceosomal RNP components Ro/SSA, and La/SSB are also targets of autoimmune-related antibody responses (White et al., 1981; Wolin and Steitz, 1984).

It is intriguing that RNP components are common targets of the autoimmune response as recent studies have shown these RNP components not only initiate autoantibody production but also perpetuate autoimmunity (Marshak-Rothstein, 2006; Martin and Elkon, 2005). These RNPs become immunogenic during apoptosis-induced cleavage, structural alteration, or interaction with autoreactive B cells. Immunoglobulin-snRNP complexes have been shown to be taken up by endocytosis through the B cell receptors on B cells or Fc-receptor on dentritic cells (Boule et al., 2004; Leadbetter et al., 2002). RNA components, particularly those rich in uridine (U) or uridine and guanosine (UG) function as effective ligands for Toll-like receptor TLR7 and TLR7/8 in B cells and dentritic cells (Diebold et al., 2004; Heil et al., 2004; Vollmer et al., 2005). TLR stimulation leads to cell activation and proliferation, cytokine production, MHCII and co-stimulatory molecule expression, thus perpetuating autoimmunity (Martin and Elkon, 2005).

Epitope tagging proteins of interest, widely used for affinity chromatography, western blotting, or immunofluorescence assays, has been a valuable tool for protein compositional, structural and functional studies (as described in the previous chapters). Interestingly, when we specifically modified the pre-mRNA splicing factor SART1 (U4/U6•U5-110K) by insertion of the TAP tag at its 3' terminus, autoimmune phenotypes were observed in mice containing a TAP tag at both loci. Only homozygotes exhibit several physical manifestations typical of mouse autoimmunity such as runting syndrome with ruffled hair, hunched back and severe anal prolapse. These mice also possess high serum IgG levels and increased numbers of MZ B cells. Histological examination showed disorganized red and white pulp centers and proliferation of lymphocytes in spleen. Infertility assays showed increased inflammatory cell infiltration and apoptosis in germ cells of homozygote testes. Additionally, the serum from some homozygote knockin mice specifically recognized a ~60kD nuclear antigen, whereas serum from normal mice were non-reactive to nuclear protein components. These mice will serve as a valuable tool for the studies of mammalian autoimmune diseases, especially those resulting from the generation of autoantibodies against RNP components, as well as serving as a biochemical tool for spliceosome analysis in live mice.

5.2 MATERIALS AND METHODS

5.2.1 Quantitation of total serum immunoglobulin concentration

Whole blood samples were collected from mice tails. After centrifugation, the supernatant was removed and stored at –80°C or immediately tested using an ELISA assay for total serum IgG or IgM using a mouse IgG or IgM ELISA Quantitation kit (Bethyl Laboratories). Microtiter plates were coated with goat anti-mouse IgG or IgM

affinity-purified Ab and incubated for 60 min. Plates were washed with [50 mM Tris-HCl (pH 8.0), 0.14 M NaCl containing 0.05% Tween 20] and each well was blocked with 200 µl of [50 mM Tris-HCl (pH 8.0), 0.14 M NaCl, and 1% BSA] for 30 min. 100 µl of test samples and standards were added per well, and incubated for 60 min. 100 µl of horseradish peroxidase (HRP)-labeled goat anti-mouse IgG or IgM Fc-specific antibodies were added to each well and incubated for 60 min. Plates were washed, developed for 30 min with HRP substrate TMB (3, 3', 5, 5'-tetramethylbenzidine), stopped with 100 µl 2 M H_2SO_4, and read at 450 nm using a microtiter plate reader.

5.2.2 Flow cytometry analysis

Spleens were harvested from euthanized mice and washed with HBSS buffer (Sigma). Tissues were passed through a 70 µm nylon cell strainer to prepare a single-cell suspension. Isolated cells from spleens were treated with RBC lysis buffer [0.5 M NH_4Cl, 0.15 M Tris-HCl, (pH 7.65)] for 5 minutes at room temperature to remove red blood cells. Cells were washed with HBSS buffer twice, harvesting at 1000 x g at 4˚C and resuspended in Hanks buffer (HBSS buffer with 2% FBS and 0.1% sodium azide) on ice. Cells were counted using a hemocytometer and $2x10^6$ cells were used for FACS analysis. Cells were incubated with Fc-blocking (kindly provided by Dr. Ellen Richie, M.D. Anderson Cancer Research center) for 15 minutes on ice. Spleen cells were stained with the following mAbs for 45 minutes: APC-conjugated anti-CD19, FITC-conjugated CD21, Bio-conjugated CD23, followed by PE-conjugated streptavidin (BD Pharmingen). The cells were then washed twice in Hanks buffer and once with HBSS buffer before fixing with 2% paraformaldehyde. Cells were analyzed on a BD FACScalibur instrument and data were analyzed using CellQuest Pro software.

5.2.3 Western blot analysis for detection of autoantibodies

Proteins from cytoplasmic and nuclear fraction were extracted from $2x10^7$ mouse STO cells. Cells were harvested by centrifugation, allowed to swell in TM buffer [10 mM Tris-HCl (pH 7.5), 3 mM $MgCl_2$] containing 0.2 mM PMSF, 1 µg/ml pepstatin, and 1 µg/ml leupeptin on ice for 10 minutes and lysed with 25 strokes of a dounce homogenizer. The nuclei were pelleted and the post-nuclear supernatant was designated as the cytoplasmic fraction. The nuclear pellets were washed twice with TM buffer containing 0.1% NP-40 and resuspended in low salt buffer [30 mM Tris-HCl (pH 7.5), 125 mM KCl, 5 mM $MgCl_2$, 0.5% Triton-X100] followed by sonication. Two pulses of 20 seconds each with 1 minute rest intervals were used at maximum output. Insoluble nuclear material was pelleted by centrifugation. For western blot analysis, 5 µg of total proteins from cytoplasmic or nuclear extracts were loaded and resolved on 10% polyacrylamide gels, transferred to nitrocellulose membranes and blotted with serum (1:1000 dilution) from the indicated mice. The secondary antibody used was HRP-conjugated goat anti-mouse IgG and the signal was detected by enhanced chemiluminescence (Perkin Elmer).

5.2.4 Histological analysis of spleen

Mice were sacrificed at 3 months of age and spleens were fixed in 10% neutral-buffered formalin. Paraffin embedded tissue sections were cut and stained with hematoxylin and eosin. Multiple sections were observed under light microscope for histological examination.

5.2.5 Immunohistochemical staining of liver samples

Mice were sacrificed at 3 months of age and livers were fixed in formalin solution. To detect the presence of SART1-TAP proteins, the slides were incubated

with PAP antibodies at a 1:50 dilution overnight at 4˚C followed by addition of peroxidase substrate DAB for staining development.

5.2.6 Histology and TUNEL assay of Testis

Mice were sacrificed at 3 months of age and testes were fixed in Bouin's solution. Sections (5 µm) of paraffin embedded tissues were evaluated for morphological changes via standard methods using periodic acid-Schiff's-Hematoxylin (PAS/H). Tissues were imaged using a Nikon E800 microscope equipped with a digital camera and captured using MetaMorph software (Downingtown, PA). Terminal deoxynucleotidyl transferase-mediated digoxigenin-dUTP nick end labeling (TUNEL assay, ApopTag^{TM} kit, Intergen, Purchase, NY) was used to detect apoptotic fragmentation of DNA in germ cells as previously described (Seaman et al., 2003). The signal from digoxigenin-dUTP end-labeled DNA was detected with anti-digoxigenin-peroxidase antibody followed by addition of peroxidase substrate [0.05% diaminobenzidine (DAB)]. TUNEL-positive germ cells were counted in each seminiferous tubule of the testis cross section. The incidence of apoptosis, represented as the apoptotic index was defined as the percentage of seminiferous tubules containing 3 or more TUNEL-positive germ cells.

5.2.7 Statistical analysis

The data were subjected to the analysis of variance (ANOVA) to determine the statistical significance. Total serum immunoglobulin concentrations and percentage of apoptosis are presented as mean ± standard error of the mean (SEM).

5.3 RESULTS

5.3.1 Tagged SART1 protein localizes in the nucleus

To demonstrate the functional expression of the TAP tag in the progeny mice, an immunohistochemistry assay was used to detect the presence of SART1-TAP using the PAP antibody. In Figure 5.1, we show that the SART1-TAP proteins were localized in the nucleus of liver cells in SART1-TAP homozygotes. As the wild-type mice did not contain SART1-TAP, no signal was detected indicating the specificity of detection in the mutant mice (Figure 5.1).

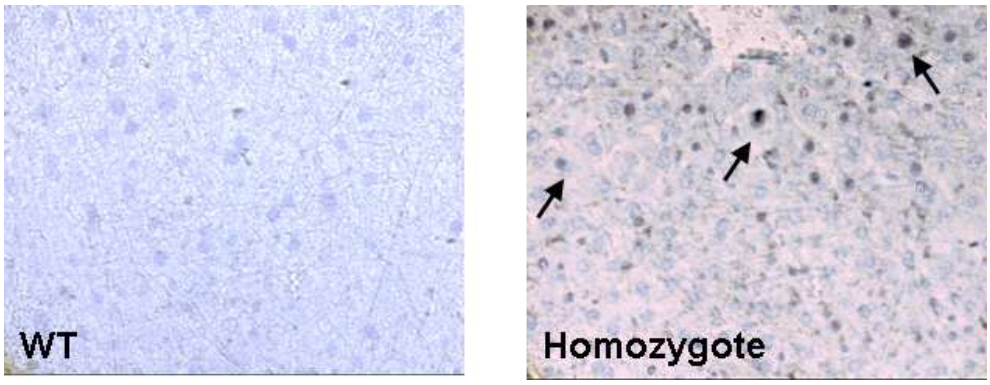

Figure 5.1 Nuclear localization of SART1-TAP in homozygote mice.

Formaldehyde fixed liver samples were sectioned, and immunohistochemistry was performed using the peroxidase-antiperoxidase (PAP) antibody as described in the Materials and Methods. Wild-type samples did not exhibit SART1 signal, while homozygote mice possess nuclear-localized SART1-TAP (arrows)

Figure 5.2 Phenotypes of SART1-TAP homozygotes.

The SART1-TAP homozygotes are smaller compared to their littermates (left panel) with ruffled hair (middle panel) and anal prolapse (right panel) symptoms.

5.3.2 Phenotypes of the SART1-TAP homozygotes

Seven out of sixteen SART1-TAP homozygotes were proportional dwarfs compared to their littermates and three died at the age of 4 months. The accompanying phenotypes include hunched back, ruffled hair, prolapsed anus and infertility (Figure 5.2). Runting syndrome is usually associated with an underlying immunological etiology, such as graft-versus-host reaction, neonatal thymectomy, cortisone injection into neonates, bacterial vaccine or endotoxin injection into neonates, and neomycin sulphate treatment (Keast, 1968). Although the severity of the clinical phenotypes in homozygotes is exhibited to different degrees in each individual, the combination of these phenotypes suggests an abnormal immunologic response in the SART1-TAP homozygotes. We feel the differences in severity are probably due to differential penetrance of the effects of the modification or individual susceptibility to the modification, similar to that seen in human autoimmunity patients.

5.3.3 Increased total serum IgG in SART1-TAP homozygotes

Total serum IgG levels were determined in wild-type, heterozygote- and homozygote-modified SART1 mice. In homozygotes, serum IgG levels were twice, on average, the IgG levels in WT mice (Figure 5.3). We note that the IgG concentration is correlated to the clinical phenotype; the more severe the phenotypes, the higher the total serum IgG concentration. This effect is not due to an overall increase in the levels of total immunoglobulin, however, as there is no significant difference in the total serum IgM concentration between the WT, heterozygotes, and homozygotes (Figure 5.4).

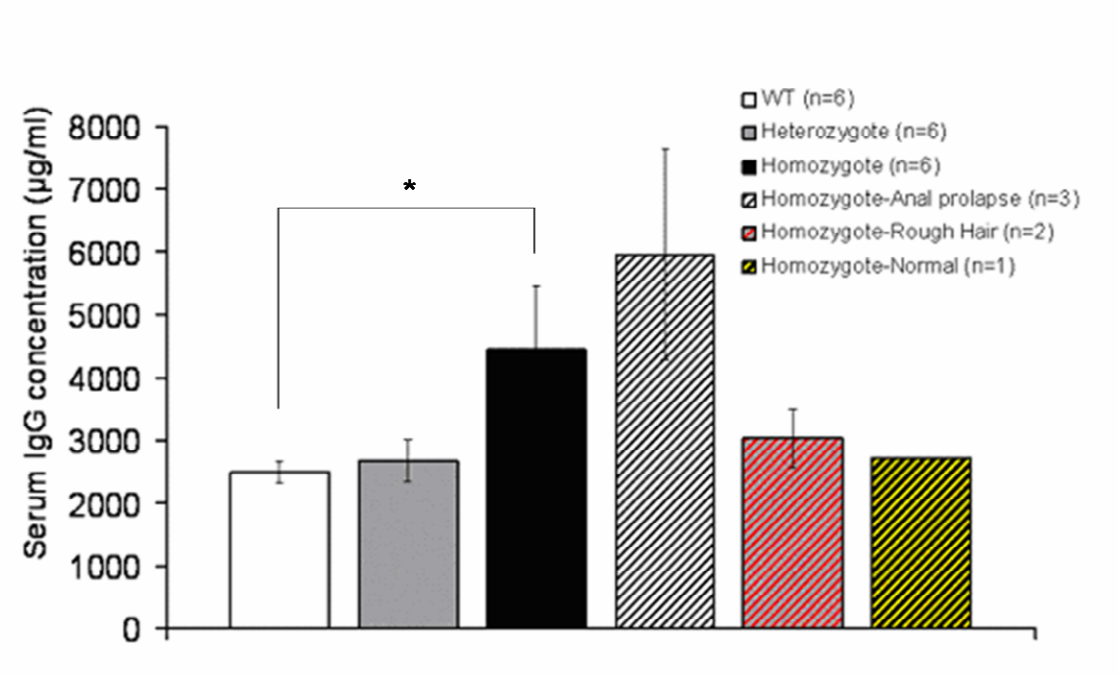

Figure 5.3 Serum IgG levels in wild-type and SART1-TAP homozygotes.

Serum IgG levels in SART1-TAP homozygote mice on average were twice that of the wild-type or SART1-TAP heterozygote mice (*P< 0.05 between WT and homozygotes). When SART1-TAP homozygote mice were compared based on the severity of the phenotypes, IgG levels were higher in mice with more severe phenotypes (cf. Anal prolapse with ruffled-hair only or no phenotype). Data are presented as ± S.E.M.

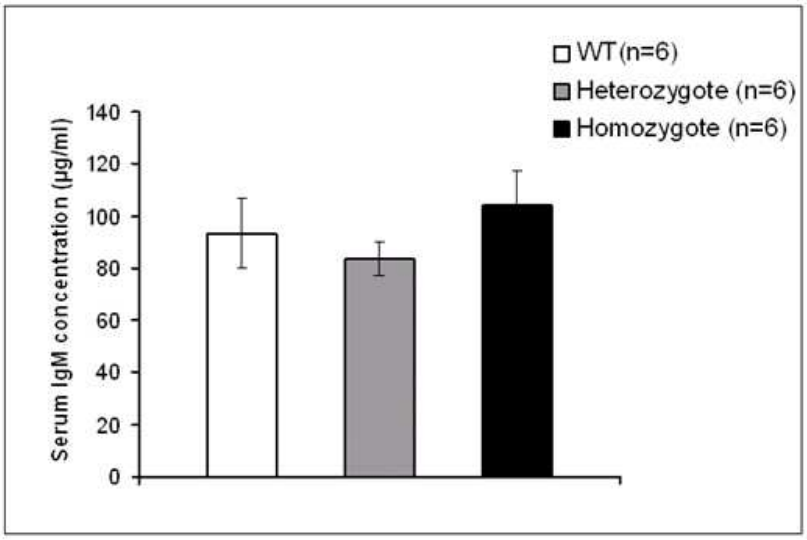

Figure 5.4 Serum IgM levels in wild-type and SART1-TAP homozygotes.

Serum IgM levels in wild-type mice were indistinguishable from that of SART1-TAP heterozygote or SART1-TAP homozygote mice. Data are presented as ± S.E.M.

5.3.4 Disorganized spleen tissue in SART1-TAP homozygotes

Typically, the organization of the spleen includes two major features termed red pulp and white pulp. The function of the red pulp is to filter out damaged red blood cells from circulation and the white pulp is a site rich in B-cells and T-cells. In wild type mice, these centers have defined boundaries (Figure 5.5). In SART1-TAP homozygotes, however, histopathological differences in spleen organization were noted, with smaller centers of red and white pulp and general disorganization (Figure 5.5).

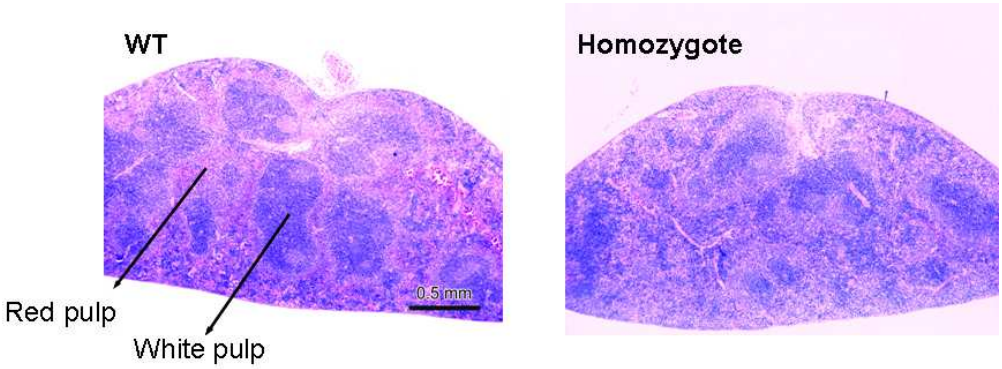

Figure 5.5 Spleen disorganization in SART1-TAP homozygotes.

Sectioned spleen from a wild-type mouse showed normal organization of red and white pulp in the left panel. Sectioned spleen from a SART1-TAP mouse showed overall histopathological disorganization and smaller centers of red and white pulp in the right panel.

5.3.5 Increased MZ B cells in SART1-TAP homozygotes

Marginal zones are regions in the spleen between the red pulp and the white pulp. Unlike circulating B cells, MZ B cells are static, and never leave this region. MZ B cells contain relatively high levels of the cell surface marker CD21 and relatively low levels of the CD23 marker protein ($CD21^{hi}CD23^{lo}$). Studies have shown MZ B cells are enriched with self-reactive clones - for example, MZ B cells are increased in the lupus-prone mice, suggesting an important role in initiation of autoimmunity. The marginal zone B cells of wild-type and SART1-TAP homozygotes were compared. In the SART1-TAP homozygote mice the MZ B cells were >2 fold higher compared to those in wild-type siblings, indicating an enrichment of self-reactive clones (Figure 5.6).

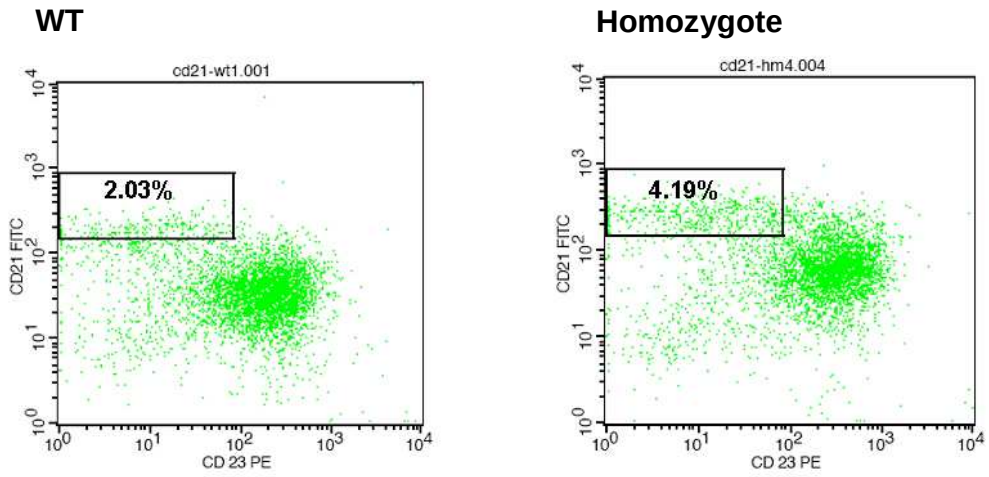

Figure 5.6 Marginal zone B-cells in wild-type and SART1-TAP homozygotes.

CD21 and CD23 were assayed in single-cell suspensions from spleens of wild-type and SART1-TAP homozygotes. Left panel: wild-type mice contain 2.03% $CD21^{hi}CD23^{lo}$ cells, indicative of the marginal zone B-cell population. Right panel: SART1-TAP homozygotes have 4.19% $CD21^{hi}CD23^{lo}$ cells, a greater than two-fold increase of this population.

5.3.6 Infertility accompanies the autoimmune phenotypes

Attempts to mate homozygote males and females were unproductive, indicating potential infertility problems in these mice. We examined the morphology and apoptosis status of the germ cells in wild-type and SART1-TAP homozygote testes. In WT mice, the germs cells aligned regularly in the seminiferous tubules and the tubules were uniform in size and shape. Cells in the center of the seminiferous tubules are more mature, with the immature cells at the periphery. In SART1-TAP homozygotes, the seminiferous tubules have an abnormal and irregular shape with various lumen sizes and the cells are aligned randomly in the tubules indicating maturation defects (Figure 5.7).

The apoptosis level in these mice was assessed by the TUNEL assay (Figure 5.8). In addition, infiltration of macrophages was observed in the interstitial tissue of the testis of homozygotes (circled in Figure 5.8). The apoptotic index was calculated as the number of seminiferous tubules with 3 or more apoptotic cells (Figure 5.9). The apoptotic level of the homozygotes was significantly higher than that of the wild type mice.

We also examined the reproductive organs of female SART1-TAP homozygotes. Female homozygotes have much smaller uteri, fallopian tubes, and ovaries compared to those in wild-type mice, suggesting they suffered from failure of normal development of these organs (Figure 5.10). These organs were harvested for histology analyses.

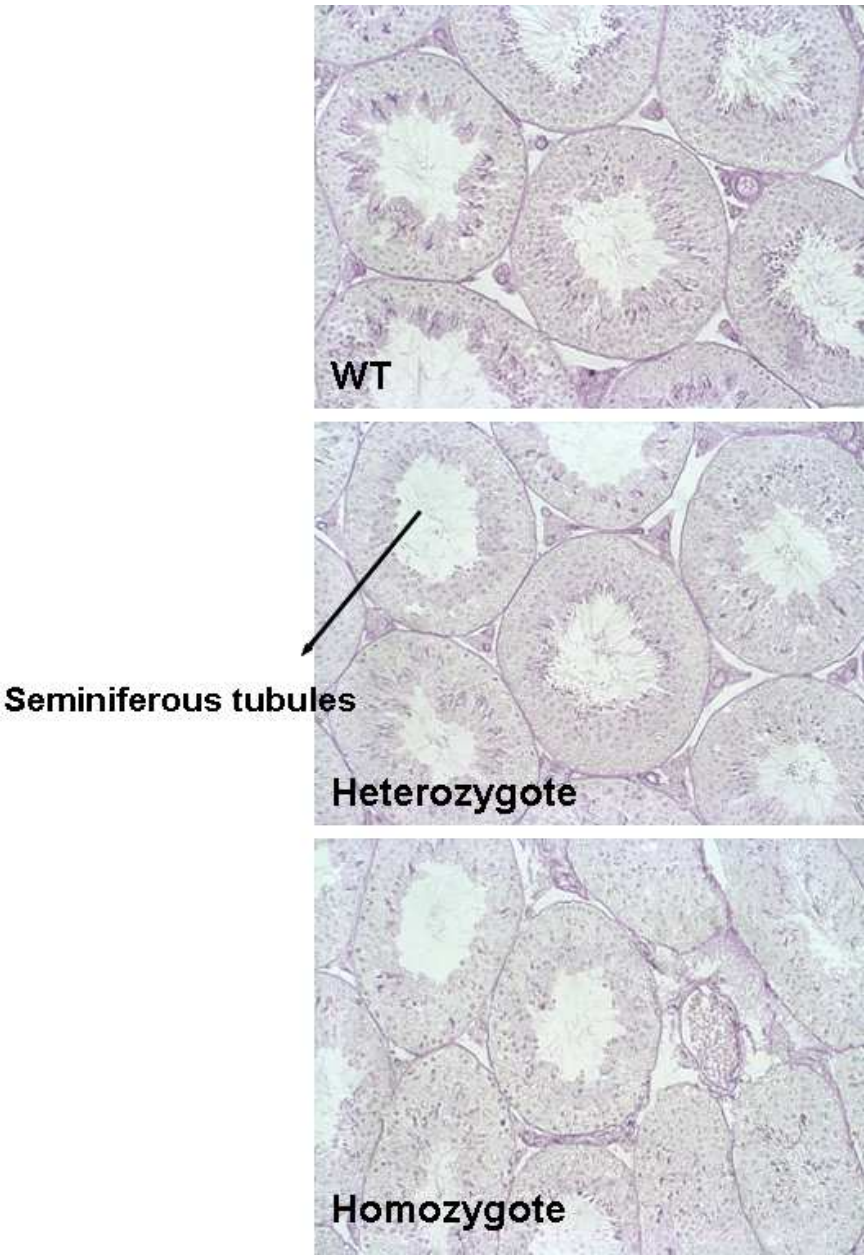

Figure 5.7 Testes histology of wild-type and SART1-TAP mice.

Sections of seminiferous tubules from wild-type (top panel), heterozygote (middle panel) or homozygote (bottom panel) mice were fixed in Bouin's solution and stained with periodic acid-Schiff's-hematoxylin. The morphology in wild-type is normal with immature cells at the periphery of the tubules and mature cells in the lumen. In contrast, the cells in the seminiferous tubules from homozygotes are disorganized with various lumen size of the tubules.

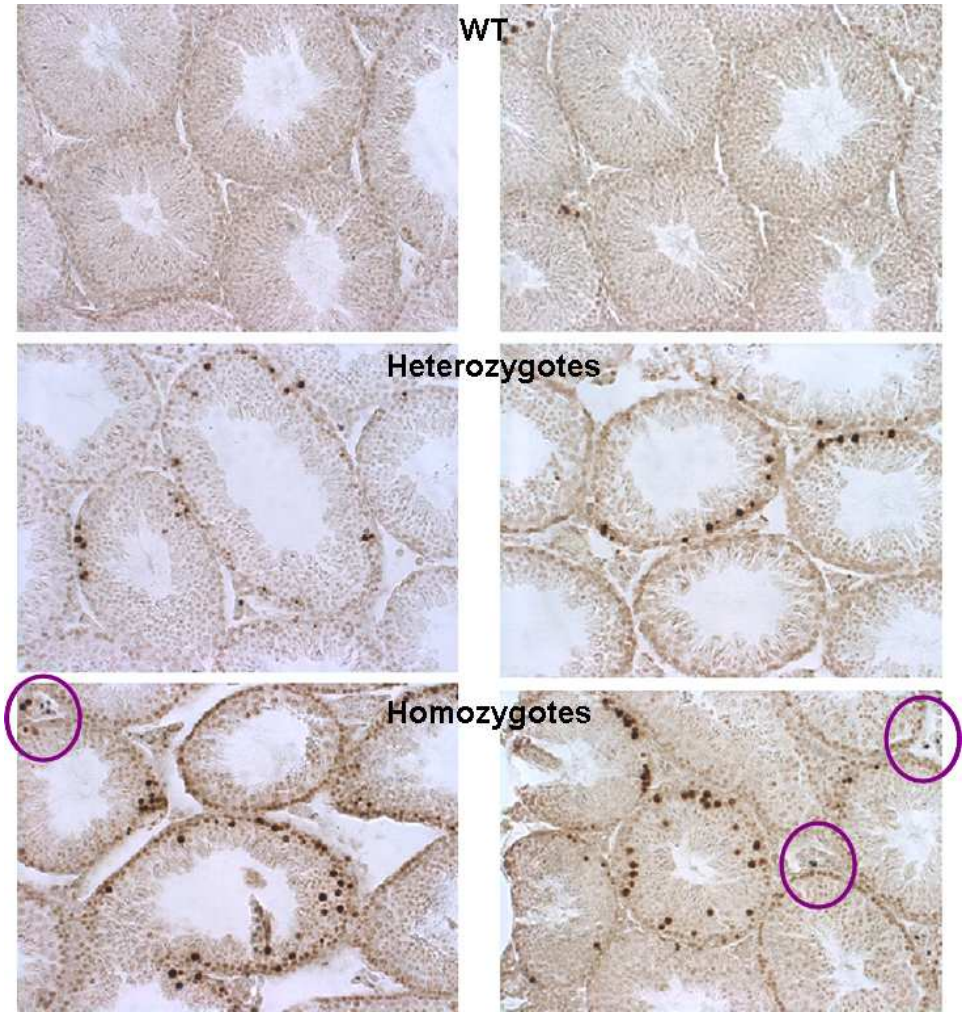

Figure 5.8 Apoptotic defects in SART1-TAP homozygotes.

Darker regions represent apoptotic cells in this assay. A gradient of apoptotic severity is seen between wild-type (top panel), heterozygote (middle panel) and homozygote (bottom panel) mice. Presumptive macrophages are shown circled and indicate likely infiltrative immune responses.

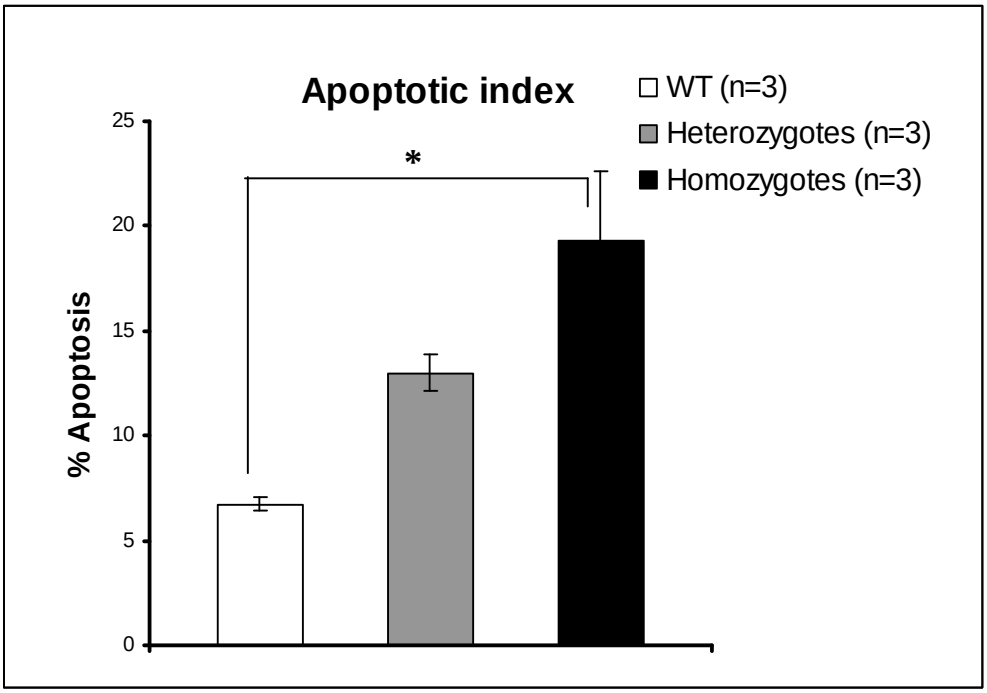

Figure 5.9 Apoptotic index in wild-type and SART1-TAP mice.

The numbers of seminiferous tubules in SART-TAP homozygotes undergo apoptosis 4 times more than those in wild-type mice (* $P < 0.05$, data are presented as ± S.E.M.).

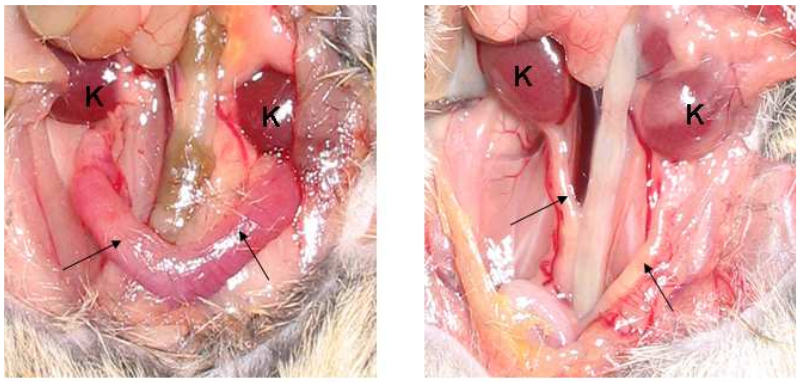

Figure 5.10 Gross view of uterus from wild-type and SART1-TAP homozygotes.

Uteri (arrows) and kidneys (**K**) were noted. Compared to the wild-type mice (left panel), the uterus in homozygotes (right panel) was much smaller, indicating failure to develop.

5.3.7 The presence of antinuclear antibodies in SART1-TAP homozygotes serum

In several forms of autoimmune disease, a spectrum of antibodies to RNP components, such as anti-Sm, anti-U1 RNP, anti-Ro/SSA, and anti-La/SSB has been seen. Using serum as a primary antibody source, we noted a signal of approximately 60Kd in the nuclear fraction in SART1-TAP homozygotes (Figure 5.11), showing that in addition to the array of autoimmune symptoms discussed above, SART1-TAP homozygotes produce anti-nuclear autoantibodies in much the same way as classical autoimmune mice do.

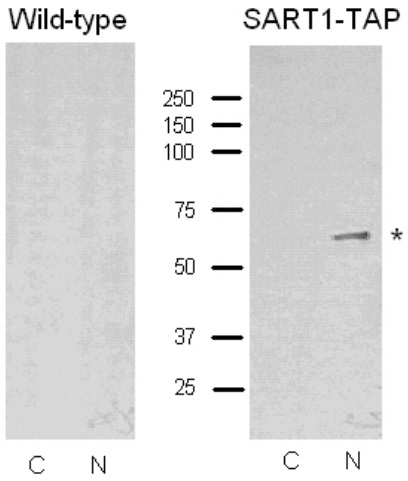

Figure 5.11 Circulating anti-nuclear antibodies in SART1-TAP homozygotes.

Serum from wild-type and SART1-TAP mice was used as a primary antibody in a western blot against total protein from the nuclear (N) or the cytoplasmic (C) fractions of cultured murine cells. After probing with a secondary, anti- mouse IgG antibody, a ~60kD antigen present exclusively in the nucleus was reactive with serum from only the SART1-TAP homozygote mice.

5.4 DISCUSSION

Autoimmune diseases are characterized by circulating autoantibodies reactive against cellular components. Among the earliest identified antibodies were anti-Sm and anti-U1 snRNP antibodies, targeting conserved nuclear components (Mattioli and Reichlin, 1971; Tan and Kunkel, 1966). Lerner and Steitz showed that anti-Sm antibodies react with the U1, U2, U4 and U5 snRNP particles whereas anti-U1 snRNP antibodies precipitate only the U1 snRNP (Lerner and Steitz, 1979). Subsequently, they

showed that the sequence of U1 RNA is complementary to 5' splice junctions in pre-mRNA, providing a link between snRNPs and pre-mRNA splicing (Lerner et al., 1980).

The question of how autoantigens stimulate inappropriate immune response has remained mysterious for several decades now. The SmB/B' protein possesses an epitope PPPGMRPP (Williams et al., 1990), located within its carboxyl-terminus, which cross-reacts with the U1 snRNP (De Keyser et al., 1992), suggesting similarities in autoimmune response generation (Arbuckle et al., 1999).

Proper folding and posttranslational modifications of the epitope have also been shown to be important for autoimmune responses. Methylation of arginine residues within the C-terminal epitope of SmD1 is critical for the recognition of SmD1 by anti-Sm antibodies (Brahms et al., 2000). In contrast, synthetic SmD1 peptides produced *in vitro* or SmD1 overexpressed in *E. coli* interact only weakly or non-specifically with anti-Sm antibodies (Rivkin et al., 1994; Rokeach et al., 1992). Interaction between anti-La/SSB antibodies and La proteins is strongly dependent on phosphorylated La protein at its serine 366 residue (Terzoglou et al., 2006a) whereas anti-Ro antibodies preferentially recognize the unmodified Ro protein (Terzoglou et al., 2006b). In addition, the antigenicity of Ro is significantly enhanced upon binding with a chaperone protein, calreticulin (Staikou et al., 2003). Although little is known about the effects of the modification of proteins on its antigenicity, modification of the SART1 protein by insertion a TAP tag at its 3' terminus may change its structure, and provoke autoimmunity.

Several studies have shown the links between apoptosis and autoimmunity (Utz et al., 2000). Novel epitopes generated during apoptosis, produced either by altered peptide structure or residue modification, are sequestered in apoptotic blebs of the cell (Casciola-Rosen et al., 1994), and are proposed to serve as targets for autoantibody

production (Casciola-Rosen et al., 1994; Hall et al., 2004; Hof et al., 2005; Utz et al., 1997). Failure of completion of apoptosis or the clearance of apoptotic cells can lead to autoimmunity by exposing novel epitopes to the immune system (Kim et al., 2003; Utz et al., 2000). Consistent with view, germ cells with a high apoptotic index are found in SART1-TAP tagged homozygotes. These apoptotic cells may result from delays in programmed cell death or failure to clear apoptotic cells. These apoptotic cells may be triggering autoantibody production in these mice.

In conclusion, modification of the SART1 protein with a TAP tag, and defects in apoptosis pathways may contribute to the autoimmune response in SART1-TAP tagged homozygotes. We would like to note that the anti-nuclear antibodies do not recognize the 130kD SART1-TAP, and thus the autoimmune response is directed against factors in the nucleus unrelated to the modification generating the response. Whether this is a common feature in autoimmune pathology remains to be seen. Our homozygote SART1-TAP mice provide a new model for studying the pathology of autoimmunity, especially those resulting from the generation of autoantibodies against RNP components.

Appendix

Appendix A Validation of matched polypeptides by the number polypeptides identified and the percentage of the polypeptide covered

		Chicken polyspliceosome		Human polyspliceosome	
ENSEMBL accession #	Polypeptide	# pep. ID	% coverage	# pep. ID	% coverage
U1 snRNP					
ENSG00000104852	U1-70K	11	38.3	4	12.1
ENSG00000077312	U1A	6	24.3		
U2 snRNP					
ENSGALG00000008038	SF3b155	35	32.2	26	28.8
ENSG00000087365	SF3b145	12	15.5	9	9.6
ENSGALG00000002531	SF3b130	23	26.1	14	14.5
ENSGALG00000000581	SF3b125	5	18	3	12.5
ENSGALG00000007937	SF3a120	21	30.1	10	21.9
ENSGALG00000021679	SF3a66	5	28.3	4	19.2
ENSGALG00000001540	SF3a60	14	29.1	9	24.4
ENSGALG00000013352	SF3b49	3	21.7		
ENSGALG00000008729	U2B"	8	29.2	4	23.1
ENSGALG00000007170	U2A'	10	43.9	8	35.3
ENSGALG00000016501	SF3b14	5	36		
ENSGALG00000020000	SF3b10	2	22.1		
U2-snRNP associated					
ENSGALG00000014395	PRP43/DDX15	25	31.6	13	19
ENSGALG00000002612	SR140	12	13.7	5	5.6
ENSGALG00000006332	SPF45	8	23.4		
ENSGALG00000008561	SPF30	3	15.3		
ENSGALG00000003824	CHERP	3	5		
U5, U4/U6 & U4/U6•U5 snRNP					
ENSGALG00000002943	U5-220K	72	33.7	41	21
ENSGALG00000003477	U5-200K	63	27.8	44	26.8
ENSGALG00000000988	U5-116K	31	34.7	22	33.3
ENSG00000175467	U4/U5•U5-110K	2	5.1	12	19.2
ENSGALG00000006001	U5-102K	31	19.5	21	27.3
ENSG00000174243	U5-100K	21	29.9	18	26.4
ENSGALG00000000465	U4/U6-90K	13	25	8	24.6
ENSG00000168883	U4/U6•U5-65K	4	12.9	6	14.8

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
ENSG00000105618	U4/U6•U5-61K	5	20.1	8	23
ENSGALG00000008857	U4/U6-60K	12	28.7	7	16.1
ENSGALG00000004874	USA-CYP	2	24.1		
ENSGALG00000000615	U5-40K	10	33.5	5	23.2
ENSGALG00000011931	U4/U6•U5-15.5k	2	12.4		
ENSGALG00000017396	U5-15K	2	16.2		
AT/AC					
ENSGALG00000005162	U11/U12-65K	2	8.6		
ENSGALG00000013005	U11/U12-48K	2	7.2		
Sm/LSM					
ENSGALG00000007250	SmB/B'	9	40.2	5	40.7
ENSGALG00000011842	SmD1	4	37.8	4	54.6
ENSG00000125743	SmD2	12	13.7	5	44.9
ENSGALG00000006596	SmD3	4	47.6	5	40.7
ENSGALG00000000137	SmE	5	79	2	19.8
ENSGALG00000011409	SmF	4	8.7	2	15.1
ENSG00000143977	SmG	3	31.3	3	35.5
ENSG00000111987	LSM2	3	39		
ENSGALG00000003385	LSM4	3	31		
ENSGALG00000009985	LSM6	3	35	3	36.3
ENSGALP00000014820	LSM8	3	52.1	2	34.3
snRNP biogenesis					
ENSGALG00000010154	SIP1	4	14.7		
ENSG00000119953	SMNrp30	2	6		
ENSGALG00000003158	Coilin	8	20.2		
SAPs					
ENSGALG00000002514	PSF	8	16.8	12	19.1
ENSGALG00000002060	TLS/FUS	10	15.7	3	13.7
ENSGALG00000012468	PRP39	4	9.8		
ENSGALG00000010501	SKIP/PRP45	16	38.8	2	10.8
ENSGALG00000016704	CDC5	24	37.7	11	16.3
ENSGALG00000005012	ISY1	4	20		
ENSGALG00000008429	CRN1	15	24.5	4	6
ENSGALG00000009257	Prp46/PRL1	15	38.6		
ENSGALG00000015061	CDC40/PRP17	15	39.3		

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
ENSGALG00000002002	SPF27	4	27	3	12.5
ENSGALG00000013919	PRP19	13	35.4	15	46
ENSGALG00000010627	PRP38	5	18.3	3	15.6
ENSGALG00000005507	p54nrb	5	14.4	5	13.2
ENSGALG00000004555	ECM2/RBM22	6	18.6		
ENSGALG00000001247	SYF2	4	8.7		
ENSGALG00000000726	ELAV/Hu	4	13.8	11	48.2
ENSGALG00000009001	CWC22	7	9.5		
ENSGALG00000008149	EWSR1 (RBP)	5	10	2	9.9
ENSGALG00000004705	BUD31	2	27.3		
ENSG00000076924	SYF1	15	21.1	13	20.9
ENSGALG00000001962	PTB	12	32.5	12	19.1
ENSGALG00000011857	LUC7/CROP	3	30.7		
ENSG00000196504	PRP40	3	4.2	2	4.5
ENSGALG00000010167	Pinin	14	23	10	16.2
ENSGALG00000012813	Prp4K	6	16.1	8	25.2
ENSGALG00000000833	RED	7	12.4	7	21.4
ENSGALG00000001500	SLU7	8	15.4		
ENSG00000063244	U2AF65	2	10.1	7	31
ENSGALG00000016198	U2AF35	2	14.6	3	16.9
ENSGALG00000001034	AAR2	5	15.1		
ENSGALG00000005525	SF2/ASF	4	20.3	3	30.6
ENSGALG00000002087	CBC80	3	3.5	12	37.7
ENSGALG00000006843	CBC20	3	28	2	15.1
ENSG00000100296	KIAA0983	5	11.3	2	6.1
ENSG00000126803	HSP70-2	5	14.8	6	10.4
ENSGALG00000006512	HSP71	22	38.8	4	6.1
ENSGALG00000009838	Aquarius	25	17.5	20	22
ENSGALG00000002014	SMU1	10	33.1	4	19.1
ENSG00000113649	CA150			4	12.1
ENSGALG00000004626	E1B-AP5	6	11.2	5	10.1
ENSGALG00000011678	DnaJ	8	8.3	4	5.6
ENSG00000100813	Acinus			10	22.6
ENSG00000131051	HCC			9	27.3
ENSG00000084463	WBP11			2	4.5
ENSG00000196419	Ku70			8	16.4
RNA Helicase-like					
ENSGALG00000016461	DDX1	10	18.6	8	16.4

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
ENSGALG00000016231	DDX3	5	10.3	9	17.4
ENSGALG00000003532	DDX5/p68	16	28.9		
ENSGALG00000012247	DDX17/p72	11	20.8	13	21.3
ENSGALG00000012147	DDX18	2	10		
ENSGALG00000006974	DDX27	3	5.1		
ENSGALG00000003030	DDX41/ABSTRAKT	3	7.4		
ENSG00000123136	DDX39			12	34.9
ENSG00000145833	DDX46/PRP5			13	19
ENSGALG00000008530	DDX48	18	48.4	15	39.7
ENSGALG00000004144	DDX50/Gu-ß	2	3.8	11	17.3
ENSGALG00000001186	DHX8/PRP22	12	17.9	2	7.7
ENSG00000135829	DHX9/HELICASEA	3	5.6	43	40.9
ENSGALG00000005027	DHX30	20	19.7	9	9.3
ENSGALG00000003658	DHX35	4	7.8		
ENSG00000174953	DHX36			5	5.1
ENSGALG00000014709	SKIV2L2	5	7.1	2	10.8
ENSG00000198563	UAP56	12	29.7	12	34.9
hnRNP					
ENSGALG00000006160	hnRNPA0	4	10.2	7	35.4
ENSGALG00000011036	hnRNPA2/B1	12	37.8	23	67.2
ENSG00000135486	hnRNPA1	7	32.7	26	52.9
ENSGALG00000009250	hnRNPA3	11	42.9	19	51.6
ENSGALG00000014381	hnRNPAB	6	20.7	7	22.1
ENSG00000092199	hnRNPC1/C2	10	32.5	14	47.9
ENSG00000138668	hnRNPD0/AUF1	4	29.7	13	45.6
ENSGALG00000011184	hnRNPD0	4	20.6	9	25.4
ENSG00000169813	hnRNPF			12	42.4
ENSGALG00000006457	hnRNPG	15	34.3	11	26.9
ENSGALG00000005955	hnRNPH1	17	46.4	13	40.3
ENSGALG00000003947	hnRNPH3	12	13.7	6	26.9
ENSGALG00000012591	hnRNPK	18	46.1	18	49.1
ENSG00000104824	hnRNPL	7	13.7	14	33
ENSGALG00000000377	hnRNPM	24	38.6	21	37.7
ENSGALG00000015830	hnRNPQ	10	27.3	18	45.8
ENSGALG00000000814	hnRNPR	11	22.5	21	42.1
ENSGALG00000010671	hnRNPU	10	13.1	29	33
ENSGALG00000018665	hnRNP novel	2	***		
ENSG00000126457	HRMT1L2	2	11.5		

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
SR					
ENSGALG00000005525	SF2p33	9	35.8	12	37.9
ENSGALG00000000533	SFRS3SRp20	3	24.4	8	59.3
ENSGALG00000009484	SRp40	5	24.8		
ENSGALG00000000990	SRp55	6	27.4	4	11.5
ENSGALG00000013825	9G8	5	26	6	37.6
ENSGALG00000002487	SFRS8	6	10.5		
ENSG00000111786	SRp30			8	37.6
ENSGALG00000006531	SFRS10	4	20.1	6	37.2
ENSG00000153914	SFRS12			3	11.9
ENSGALG00000004133	SRrp35	3	37.1	6	41.8
RBP					
ENSG00000102317	RBM3			25	5.6
ENSGALG00000018992	RBM4B/Lark	9	56.8		
ENSGALG00000004626	RBM5	4	9.1		
ENSGALG00000007017	RBM7	3	36		
ENSG00000131795	RBM8B	3	17.7	4	31
ENSGALG00000013411	RBM14	6	12.9	8	15.7
ENSG00000162775	RBM15			7	12
ENSGALG00000002372	RBM15B	7	15.6	3	8.5
ENSG00000197676	RBM16			8	12.3
ENSGALG00000004555	RBM22	5	17.4		
ENSG00000119707	RBM25			4	8.4
ENSG00000091009	RBM27			2	3.2
ENSGALG00000011038	RNPS1	3	17		
ENSG00000033030	ZFP8			6	8.1
ENSGALG00000006113	ZFP326	10	23.2	2	9.4
ENSG00000179950	PUF60			3	10.9
ENSG00000197381	ADAR1			26	33.7
ENSG00000160710	ADAR2			4	10
ENSGALG00000010952	Requiem	12	27.2		
ENSGALG00000011570	NFAT45	8	21.4	12	53.7
ENSG00000129351	NFAT90	10	15.5	12	18.4
ENSG00000169564	PolyrCBP	7	28.7		
ENSGALG00000012225	CIRBP	2	24.5	3	18.9
ENSG00000056097	ZFR	2	3.4	3	8
ENSGALG00000009427	TIA1	2	8.8		
ENSGALG00000001187	STRBP	10	18.5		

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
ENSGALG00000005623	Tuftillin-IP	8	12.4	5	6.7
ENSG00000136231	IMP3	4	7.5	7	18.9
ENSG00000121774	SAM68			4	10.2
ENSG00000126254	Novel RRM			3	14
ENSG00000132773	TOE1			2	6.7
ENSG00000142864	SERBP1			2	9.1
Export/transcription/NMD					
ENSGALG00000014915	THO1/HPR1	11	26	6	14.6
ENSGALG00000008507	THO2	14	9.7	5	8
ENSG00000051596	THO3/TEX1	14	24.2	13	16.3
ENSGALG00000007237	THO4/ALY	6	20.9	8	30.8
ENSGALG00000004571	GLC7/PPP1CA	2	5.3		
ENSGALG00000002569	Ran	5	30.1	2	4.8
ENSG00000119392	GLE1L			2	5.7
ENSGALG00000007653	GLE2/RAE1	3	8.4	7	13.1
ENSGALG00000003220	UPF1/RENT	3	6.4		
ENSG00000131795	Y14/RBM8A	4	36.2	4	31
ENSGALG00000002144	TRAP150	9	20.9	3	24.3
ENSG00000172660	TAF15/RBP56	4	9.2		
ENSG00000162231	TAP			7	13.1
ENSG00000065978	YBX1			5	33.5
ENSGALG00000010689	Mago nashi	7	52.8	7	57.9
3' end proc.					
ENSGALG00000010783	CPSF2	2	10.5		
ENSGALG00000016424	CPSF3	2	8.1	2	4.8
ENSGALG00000004714	CPSF4	4	20		
ENSGALG00000003084	CPSF5			4	21.6
ENSGALG00000011685	CSTF-77	3	4.5	3	6.2
ENSGALG00000013943	FIP1	4	7.8		
ENSG00000172239	PAIP1			5	14.2
ENSG00000100836	PABPN1			5	16.9
ENSGALG00000003800	PABPC4			10	22.6
Structural					
ENSGALG00000012533	Myosin	44	28.4	26	32.8
ENSGALG00000009126	Titin	12	***		
ENSGALG00000001381	Actin	15	52	3	17.6

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
ENSGALG00000002478	MATRIN3	17	24.2	28	34.2
ENSGALG00000008677	Vimentin	13	40.1	24	50.7
ENSG00000117245	Kinesin KIF17	7	8		
ENSGALG00000002197	NUMATRIN	4	17	4	4.5
ENSGALG00000014692	Lamin B	18	35.8	4	9.9
ENSGALG00000013505	NuSpectrin	12	***	34	19.9
Chromatin modification					
ENSGALG00000000360	ARID1A-SWI/SNF	18	27.4		
ENSGALG00000013683	ARID1B-SWI/SNF	13	9.8		
ENSGALG00000010164	SMARCA2	27	33.4		
ENSG00000127616	Brahma/SMARCA4	29	21.3		
ENSGALG00000009913	SMARCA5	11	11		
ENSGALG00000005983	SMARCB1	11	37		
ENSGALG00000005048	BRG1-SWI/SNF	30	20.2		
ENSGALG00000005048	SMARCC1	21	19.5		
ENSGALP00000010010	SMARCD1	6	11.5		
ENSGALG00000000363	SMARCD2	10	32.9		
ENSGALG00000002100	SMARCE1	7	30.9		
Nucleoporins					
ENSGALG00000005714	PKD1 (NUP assoc)	10	4.1		
ENSGALG00000003830	NUP214	7	4.8	4	4
ENSGALG00000012720	NUP153	4	4.3		
ENSGALG00000005078	NUP210	6	4.9	8	5.1
ENSG00000102900	NUP93			19	29.4
ENSG00000108559	NUP88			8	13.9
ENSG00000111581	NUP107			5	31.3
ENSG00000110713	NUP98			4	11.9
ENSG00000138750	NUP54			4	21.3
ENSG00000163002	NUP35			3	14.6
ENSG00000069248	NUP133			3	3.4
ENSG00000155561	NUP205			4	3.3
Cyclophillins					
ENSGALG00000013383	CYP-E	4	20.4		
ENSGALG00000004874	USA-CYP	2	34.1		
ENSGALG00000014747	CYP16	4	12.4		
ENSG00000137168	PPIL1/CWF27	2	12.2		

(Continued)

ENSEMBL accession #	Polypeptide	Chicken polyspliceosome # pep. ID	Chicken polyspliceosome % coverage	Human polyspliceosome # pep. ID	Human polyspliceosome % coverage
ENSG00000100023	PPIL2/CYP60	2	6.1		
ENSG00000113593	PPWD1	4	7.4		
Novel or unknown to splicing					
ENSGALG00000011351	HSP90a	11	18.2	10	19.3
ENSGALG00000010175	HSP90ß	18	23.4	9	14.6
ENSGALG00000012726	HSP108	15	19	8	14.5
ENSGALG00000009967	LRP130	19	13.5		
ENSGALG00000003693	Macrophin	6	16.9		
ENSGALG00000007705	Nucleolin	10	16.4		
ENSGALG00000015933	GCRBF	3	5.9		
ENSGALG00000008454	NOP58	5	12.5		
ENSGALG00000009061	BAF53A	9	32		
ENSGALG00000007520	FACT80	9	13.8	3	5.6
ENSGALG00000015821	TCP1-theta	7	15.9		
ENSGALG00000014500	Nol5A/NOP56	9	19.4		
ENSGALG00000005624	TRAP230	2	4.6		
ENSGALG00000010973	TRA2a	4	20	2	16.8
ENSGALG00000008372	RAT1	8	10	3	3.7
ENSG00000197157	SND1	5	8.8	16	25.9
ENSGALG00000001948	SAFB/HSP27	3	8.2	12	78.9
ENSGALG00000003177	BRD8	4	***	2	6.9
ENSGALG00000010699	FOG	3	***		
ENSGALG00000005177	C9ORF10	4	6.8	3	5
ENSGALG00000002653	ELG	4	13.2	3	8.7
ENSGALG00000016949	WBP4	5	14.4		
ENSGALG00000004133	Fus IP	3	37.1	6	41.8
ENSGALG00000017384	ERH	3	65.6	3	33.3
ENSG00000079246	Ku80			14	28.1
ENSG00000182562	ATAD3A	12	24.4	9	18.6
ENSGALG00000001515	ATAD3B	10	19	6	21.3
ENSG00000108588	CCDC47			2	18.4

References

Abovich, N., P. Legrain, and M. Rosbash. 1990. The yeast PRP6 gene encodes a U4/U6 small nuclear ribonucleoprotein particle (snRNP) protein, and the PRP9 gene encodes a protein required for U2 snRNP binding. *Mol Cell Biol*, 10:6417-25.

Achsel, T., H. Brahms, B. Kastner, A. Bachi, M. Wilm, and R. Luhrmann. 1999. A doughnut-shaped heteromer of human Sm-like proteins binds to the 3'-end of U6 snRNA, thereby facilitating U4/U6 duplex formation in vitro. *Embo J*, 18:5789-802.

Albers, M., A. Diment, M. Muraru, C. S. Russell, and J. D. Beggs. 2003. Identification and characterization of Prp45p and Prp46p, essential pre-mRNA splicing factors. *Rna*, 9:138-50.

Amberg, D. C., A. L. Goldstein, and C. N. Cole. 1992. Isolation and characterization of RAT1: an essential gene of Saccharomyces cerevisiae required for the efficient nucleocytoplasmic trafficking of mRNA. *Genes Dev*, 6:1173-89.

Arbuckle, M. R., M. Reichlin, J. B. Harley, and J. A. James. 1999. Shared early autoantibody recognition events in the development of anti-Sm B/B' in human lupus. *Scand J Immunol*, 50:447-55.

Arenas, J. E., and J. N. Abelson. 1997. Prp43: An RNA helicase-like factor involved in spliceosome disassembly. *Proc Natl Acad Sci U S A*, 94:11798-802.

Asano, M., M. Toda, N. Sakaguchi, and S. Sakaguchi. 1996. Autoimmune disease as a consequence of developmental abnormality of a T cell subpopulation. *J Exp Med*, 184:387-96.

Azubel, M., N. Habib, R. Sperling, and J. Sperling. 2006. Native spliceosomes assemble with pre-mRNA to form supraspliceosomes. *J Mol Biol*, 356:955-66.

Azubel, M., S. G. Wolf, J. Sperling, and R. Sperling. 2004. Three-dimensional structure of the native spliceosome by cryo-electron microscopy. *Mol Cell*, 15:833-9.

Baba, T. W., B. P. Giroir, and E. H. Humphries. 1985. Cell lines derived from avian lymphomas exhibit two distinct phenotypes. *Virology*, 144:139-51.

Bach, M., G. Winkelmann, and R. Luhrmann. 1989. 20S small nuclear ribonucleoprotein U5 shows a surprisingly complex protein composition. *Proc Natl Acad Sci U S A*, 86:6038-42.

Banroques, J., and J. N. Abelson. 1989. PRP4: a protein of the yeast U4/U6 small nuclear ribonucleoprotein particle. *Mol Cell Biol*, 9:3710-9.

Batsche, E., M. Yaniv, and C. Muchardt. 2006. The human SWI/SNF subunit Brm is a regulator of alternative splicing. *Nat Struct Mol Biol*, 13:22-9.

Beggs, J. D. 1978. Transformation of yeast by a replicating hybrid plasmid. *Nature*, 275:104-9.

Behrens, S. E., F. Galisson, P. Legrain, and R. Luhrmann. 1993a. Evidence that the 60-kDa protein of 17S U2 small nuclear ribonucleoprotein is immunologically and functionally related to the yeast PRP9 splicing factor and is required for the efficient formation of prespliceosomes. *Proc Natl Acad Sci U S A*, 90:8229-33.

Behrens, S. E., and R. Luhrmann. 1991. Immunoaffinity purification of a [U4/U6.U5] tri-snRNP from human cells. *Genes Dev*, 5:1439-52.

Behrens, S. E., K. Tyc, B. Kastner, J. Reichelt, and R. Luhrmann. 1993b. Small nuclear ribonucleoprotein (RNP) U2 contains numerous additional proteins and has a bipartite RNP structure under splicing conditions. *Mol Cell Biol*, 13:307-19.

Ben-Yehuda, S., I. Dix, C. S. Russell, M. McGarvey, J. D. Beggs, and M. Kupiec. 2000. Genetic and physical interactions between factors involved in both cell cycle progression and pre-mRNA splicing in Saccharomyces cerevisiae. *Genetics*, 156:1503-17.

Bennett, M., S. Michaud, J. Kingston, and R. Reed. 1992. Protein components specifically associated with prespliceosome and spliceosome complexes. *Genes Dev*, 6:1986-2000.

Berget, S. M. 1995. Exon recognition in vertebrate splicing. *J Biol Chem*, 270:2411-4.

Berget, S. M., C. Moore, and P. A. Sharp. 1977. Spliced segments at the 5' terminus of adenovirus 2 late mRNA. *Proc Natl Acad Sci U S A*, 74:3171-5.

Berglund, J. A., K. Chua, N. Abovich, R. Reed, and M. Rosbash. 1997. The splicing factor BBP interacts specifically with the pre-mRNA branchpoint sequence UACUAAC. *Cell*, 89:781-7.

Blanton, S., A. Srinivasan, and B. C. Rymond. 1992. PRP38 encodes a yeast protein required for pre-mRNA splicing and maintenance of stable U6 small nuclear RNA levels. *Mol Cell Biol*, 12:3939-47.

Blevins, M. B., A. M. Smith, E. M. Phillips, and M. A. Powers. 2003. Complex formation among the RNA export proteins Nup98, Rae1/Gle2, and TAP. *J Biol Chem*, 278:20979-88.

Boule, M. W., C. Broughton, F. Mackay, S. Akira, A. Marshak-Rothstein, and I. R. Rifkin. 2004. Toll-like receptor 9-dependent and -independent dendritic cell activation by chromatin-immunoglobulin G complexes. *J Exp Med*, 199:1631-40.

Bourgeois, C. F., F. Lejeune, and J. Stevenin. 2004. Broad specificity of SR (serine/arginine) proteins in the regulation of alternative splicing of pre-messenger RNA. *Prog Nucleic Acid Res Mol Biol*, 78:37-88.

Bradley, A., M. Evans, M. H. Kaufman, and E. Robertson. 1984. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature*, 309:255-6.

Brahms, H., J. Raymackers, A. Union, F. de Keyser, L. Meheus, and R. Luhrmann. 2000. The C-terminal RG dipeptide repeats of the spliceosomal Sm proteins D1 and D3 contain symmetrical dimethylarginines, which form a major B-cell epitope for anti-Sm autoantibodies. *J Biol Chem*, 275:17122-9.

Bringmann, P., and R. Luhrmann. 1986. Purification of the individual snRNPs U1, U2, U5 and U4/U6 from HeLa cells and characterization of their protein constituents. *Embo J*, 5:3509-16.

Bringmann, P., J. Rinke, B. Appel, R. Reuter, and R. Luhrmann. 1983. Purification of snRNPs U1, U2, U4, U5 and U6 with 2,2,7-trimethylguanosine-specific antibody and definition of their constituent proteins reacting with anti-Sm and anti-(U1)RNP antisera. *Embo J*, 2:1129-35.

Buerstedde, J. M., and S. Takeda. 1991. Increased ratio of targeted to random integration after transfection of chicken B cell lines. *Cell*, 67:179-88.

Busch, H., R. Reddy, L. Rothblum, and Y. C. Choi. 1982. SnRNAs, SnRNPs, and RNA processing. *Annu Rev Biochem*, 51:617-54.

Casciola-Rosen, L. A., G. Anhalt, and A. Rosen. 1994. Autoantigens targeted in systemic lupus erythematosus are clustered in two populations of surface structures on apoptotic keratinocytes. *J Exp Med*, 179:1317-30.

Caspary, F., A. Shevchenko, M. Wilm, and B. Seraphin. 1999. Partial purification of the yeast U2 snRNP reveals a novel yeast pre-mRNA splicing factor required for pre-spliceosome assembly. *Embo J*, 18:3463-74.

Chan, S. P., D. I. Kao, W. Y. Tsai, and S. C. Cheng. 2003. The Prp19p-associated complex in spliceosome activation. *Science*, 302:279-82.

Chen, J. H., and R. J. Lin. 1990. The yeast PRP2 protein, a putative RNA-dependent ATPase, shares extensive sequence homology with two other pre-mRNA splicing factors. *Nucleic Acids Res*, 18:6447.

Chen, Y. I., S. D. Maika, and S. W. Stevens. 2006. Epitope tagging of proteins at the native chromosomal loci of genes in mice and in cultured vertebrate cells. *J Mol Biol*, 361:412-9.

Cheng, S. C., and J. Abelson. 1987. Spliceosome assembly in yeast. *Genes Dev*, 1:1014-27.

Cheng, S. C., W. Y. Tarn, T. Y. Tsao, and J. Abelson. 1993. PRP19: a novel spliceosomal component. *Mol Cell Biol*, 13:1876-82.

Cho, E. J., M. S. Kobor, M. Kim, J. Greenblatt, and S. Buratowski. 2001. Opposing effects of Ctk1 kinase and Fcp1 phosphatase at Ser 2 of the RNA polymerase II C-terminal domain. *Genes Dev*, 15:3319-29.

Chow, L. T., R. E. Gelinas, T. R. Broker, and R. J. Roberts. 1977. An amazing sequence arrangement at the 5' ends of adenovirus 2 messenger RNA. *Cell*, 12:1-8.

Chung, S., M. R. McLean, and B. C. Rymond. 1999. Yeast ortholog of the Drosophila crooked neck protein promotes spliceosome assembly through stable U4/U6.U5 snRNP addition. *Rna*, 5:1042-54.

Colot, H. V., F. Stutz, and M. Rosbash. 1996. The yeast splicing factor Mud13p is a commitment complex component and corresponds to CBP20, the small subunit of the nuclear cap-binding complex. *Genes Dev*, 10:1699-708.

Company, M., J. Arenas, and J. Abelson. 1991. Requirement of the RNA helicase-like protein PRP22 for release of messenger RNA from spliceosomes. *Nature*, 349:487-93.

Cramer, P., J. F. Caceres, D. Cazalla, S. Kadener, A. F. Muro, F. E. Baralle, and A. R. Kornblihtt. 1999. Coupling of transcription with alternative splicing: RNA pol II promoters modulate SF2/ASF and 9G8 effects on an exonic splicing enhancer. *Mol Cell*, 4:251-8.

Dalbadie-McFarland, G., and J. Abelson. 1990. PRP5: a helicase-like protein required for mRNA splicing in yeast. *Proc Natl Acad Sci U S A*, 87:4236-40.

Davis, M. T., and T. D. Lee. 1998. Rapid protein identification using a microscale electrospray LC/MS system on an ion trap mass spectrometer. *J Am Soc Mass Spectrom*, 9:194-201.

De Keyser, F., S. O. Hoch, M. Takei, H. Dang, H. De Keyser, L. A. Rokeach, and N. Talal. 1992. Cross-reactivity of the B/B' subunit of the Sm ribonucleoprotein autoantigen with proline-rich polypeptides. *Clin Immunol Immunopathol*, 62:285-90.

de Lanerolle, P., T. Johnson, and W. A. Hofmann. 2005. Actin and myosin I in the nucleus: what next? *Nat Struct Mol Biol*, 12:742-6.
Dellaire, G., E. M. Makarov, J. J. Cowger, D. Longman, H. G. Sutherland, R. Luhrmann, J. Torchia, and W. A. Bickmore. 2002. Mammalian PRP4 kinase copurifies and interacts with components of both the U5 snRNP and the N-CoR deacetylase complexes. *Mol Cell Biol*, 22:5141-56.
Diebold, S. S., T. Kaisho, H. Hemmi, S. Akira, and C. Reis e Sousa. 2004. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. *Science*, 303:1529-31.
Dix, I., C. Russell, S. B. Yehuda, M. Kupiec, and J. D. Beggs. 1999. The identification and characterization of a novel splicing protein, Isy1p, of Saccharomyces cerevisiae. *Rna*, 5:360-8.
Doetschman, T., R. G. Gregg, N. Maeda, M. L. Hooper, D. W. Melton, S. Thompson, and O. Smithies. 1987. Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature*, 330:576-8.
Eliceiri, G. L. 1980. Formation of low molecular weight RNA species in HeLa cells. *J Cell Physiol*, 102:199-207.
Epstein, P., R. Reddy, D. Henning, and H. Busch. 1980. The nucleotide sequence of nuclear U6 (4.7 S) RNA. *J Biol Chem*, 255:8901-6.
Feederle, R., H. J. Delecluse, J. P. Rouault, A. Schepers, and W. Hammerschmidt. 2004. Efficient somatic gene targeting in the lymphoid human cell line DG75. *Gene*, 343:91-7.
Figueroa, J. D., and M. J. Hayman. 2004. The human Ski-interacting protein functionally substitutes for the yeast PRP45 gene. *Biochem Biophys Res Commun*, 319:1105-9.
Fong, N., and D. L. Bentley. 2001. Capping, splicing, and 3' processing are independently stimulated by RNA polymerase II: different functions for different segments of the CTD. *Genes Dev*, 15:1783-95.
Fong, Y. W., and Q. Zhou. 2001. Stimulatory effect of splicing factors on transcriptional elongation. *Nature*, 414:929-33.
Fontenot, J. D., M. A. Gavin, and A. Y. Rudensky. 2003. Foxp3 programs the development and function of CD4+CD25+ regulatory T cells. *Nat Immunol*, 4:330-6.
Forch, P., L. Merendino, C. Martinez, and J. Valcarcel. 2003. U2 small nuclear ribonucleoprotein particle (snRNP) auxiliary factor of 65 kDa, U2AF65, can promote U1 snRNP recruitment to 5' splice sites. *Biochem J*, 372:235-40.
Fortes, P., D. Bilbao-Cortes, M. Fornerod, G. Rigaut, W. Raymond, B. Seraphin, and I. W. Mattaj. 1999. Luc7p, a novel yeast U1 snRNP protein with a role in 5' splice site recognition. *Genes Dev*, 13:2425-38.
Fouraux, M. A., M. J. Kolkman, A. Van der Heijden, A. S. De Jong, W. J. Van Venrooij, and G. J. Pruijn. 2002. The human La (SS-B) autoantigen interacts with DDX15/hPrp43, a putative DEAH-box RNA helicase. *Rna*, 8:1428-43.
Gall, J. G., M. Bellini, Z. Wu, and C. Murphy. 1999. Assembly of the nuclear transcription and processing machinery: Cajal bodies (coiled bodies) and transcriptosomes. *Mol Biol Cell*, 10:4385-402.

Gerbi, S. A., A. V. Borovjagin, and T. S. Lange. 2003. The nucleolus: a site of ribonucleoprotein maturation. *Curr Opin Cell Biol*, 15:318-25.
Ghaemmaghami, S., W. K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman. 2003. Global analysis of protein expression in yeast. *Nature*, 425:737-41.
Gil, A., P. A. Sharp, S. F. Jamison, and M. A. Garcia-Blanco. 1991. Characterization of cDNAs encoding the polypyrimidine tract-binding protein. *Genes Dev*, 5:1224-36.
Gilbert, W., and C. Guthrie. 2004. The Glc7p nuclear phosphatase promotes mRNA export by facilitating association of Mex67p with mRNA. *Mol Cell*, 13:201-12.
Goldstrohm, A. C., A. L. Greenleaf, and M. A. Garcia-Blanco. 2001. Co-transcriptional splicing of pre-messenger RNAs: considerations for the mechanism of alternative splicing. *Gene*, 277:31-47.
Gottschalk, A., B. Kastner, R. Luhrmann, and P. Fabrizio. 2001. The yeast U5 snRNP coisolated with the U1 snRNP has an unexpected protein composition and includes the splicing factor Aar2p. *Rna*, 7:1554-65.
Gottschalk, A., G. Neubauer, J. Banroques, M. Mann, R. Luhrmann, and P. Fabrizio. 1999. Identification by mass spectrometry and functional analysis of novel proteins of the yeast [U4/U6.U5] tri-snRNP. *Embo J*, 18:4535-48.
Gottschalk, A., J. Tang, O. Puig, J. Salgado, G. Neubauer, H. V. Colot, M. Mann, B. Seraphin, M. Rosbash, R. Luhrmann, and P. Fabrizio. 1998. A comprehensive biochemical and genetic analysis of the yeast U1 snRNP reveals five novel proteins. *Rna*, 4:374-93.
Gotzmann, J., C. Gerner, M. Meissner, K. Holzmann, R. Grimm, W. Mikulits, and G. Sauermann. 2000. hNMP 200: a novel human common nuclear matrix protein combining structural and regulatory functions. *Exp Cell Res*, 261:166-79.
Grabowski, P. J. 2005. Splicing-active nuclear extracts from rat brain. *Methods*, 37:323-30.
Graham, D. R., C. P. Garnham, Q. Fu, J. Robbins, and J. E. Van Eyk. 2005. Improvements in two-dimensional gel electrophoresis by utilizing a low cost "in-house" neutral pH sodium dodecyl sulfate-polyacrylamide gel electrophoresis system. *Proteomics*, 5:2309-14.
Grainger, R. J., and J. D. Beggs. 2005. Prp8 protein: at the heart of the spliceosome. *Rna*, 11:533-57.
Graveley, B. R. 2000. Sorting out the complexity of SR protein functions. *Rna*, 6:1197-211.
Gregersen, P. K., and T. W. Behrens. 2006. Genetics of autoimmune diseases--disorders of immune homeostasis. *Nat Rev Genet*, 7:917-28.
Guialis, A., M. Moraitou, M. Patrinou-Georgoula, and A. Dangli. 1991. A novel 40S multi-snRNP complex isolated from rat liver nuclei. *Nucleic Acids Res*, 19:287-96.
Guo, T. B., L. G. Boros, K. C. Chan, A. P. Hikim, A. P. Hudson, R. S. Swerdloff, A. P. Mitchell, and W. A. Salameh. 2003. Spermatogenetic expression of RNA-binding motif protein 7, a protein that interacts with splicing factors. *J Androl*, 24:204-14.

Hall, J. C., L. Casciola-Rosen, and A. Rosen. 2004. Altered structure of autoantigens during apoptosis. *Rheum Dis Clin North Am*, 30:455-71, vii.
Heath, V. L., A. Saoudi, B. P. Seddon, N. C. Moore, D. J. Fowell, and D. W. Mason. 1996. The role of the thymus in the control of autoimmunity. *J Autoimmun*, 9:241-6.
Heil, F., H. Hemmi, H. Hochrein, F. Ampenberger, C. Kirschning, S. Akira, G. Lipford, H. Wagner, and S. Bauer. 2004. Species-specific recognition of single-stranded RNA via toll-like receptor 7 and 8. *Science*, 303:1526-9.
Hengartner, C. J., V. E. Myer, S. M. Liao, C. J. Wilson, S. S. Koh, and R. A. Young. 1998. Temporal regulation of RNA polymerase II by Srb10 and Kin28 cyclin-dependent kinases. *Mol Cell*, 2:43-53.
Hinnen, A., J. B. Hicks, and G. R. Fink. 1978. Transformation of yeast. *Proc Natl Acad Sci U S A*, 75:1929-33.
Hinterberger, M., I. Pettersson, and J. A. Steitz. 1983. Isolation of small nuclear ribonucleoproteins containing U1, U2, U4, U5, and U6 RNAs. *J Biol Chem*, 258:2604-13.
Hodnett, J. L., and H. Busch. 1968. Isolation and characterization of uridylic acid-rich 7 S ribonucleic acid of rat liver nuclei. *J Biol Chem*, 243:6334-42.
Hoet, R. M., B. Kastner, R. Luhrmann, and W. J. van Venrooij. 1993. Purification and characterization of human autoantibodies directed to specific regions on U1RNA; recognition of native U1RNP complexes. *Nucleic Acids Res*, 21:5130-6.
Hof, D., J. M. Raats, and G. J. Pruijn. 2005. Apoptotic modifications affect the autoreactivity of the U1 snRNP autoantigen. *Autoimmun Rev*, 4:380-8.
Hogan, B. 1994. *Manipulating the mouse embryo : a laboratory manual*. 2nd ed. Cold Spring Harbor Laboratory Press, [Plainview, N.Y.].
Horowitz, D. S., and J. Abelson. 1993. A U5 small nuclear ribonucleoprotein particle protein involved only in the second step of pre-mRNA splicing in Saccharomyces cerevisiae. *Mol Cell Biol*, 13:2959-70.
Horowitz, D. S., R. Kobayashi, and A. R. Krainer. 1997. A new cyclophilin and the human homologues of yeast Prp3 and Prp4 form a complex associated with U4/U6 snRNPs. *Rna*, 3:1374-87.
Howe, K. J. 2002. RNA polymerase II conducts a symphony of pre-mRNA processing activities. *Biochim Biophys Acta*, 1577:308-24.
Huh, W. K., J. V. Falvo, L. C. Gerke, A. S. Carroll, R. W. Howson, J. S. Weissman, and E. K. O'Shea. 2003. Global analysis of protein localization in budding yeast. *Nature*, 425:686-91.
Jady, B. E., X. Darzacq, K. E. Tucker, A. G. Matera, E. Bertrand, and T. Kiss. 2003. Modification of Sm small nuclear RNAs occurs in the nucleoplasmic Cajal body following import from the cytoplasm. *Embo J*, 22:1878-88.
Jamison, S. F., Z. Pasman, J. Wang, C. Will, R. Luhrmann, J. L. Manley, and M. A. Garcia-Blanco. 1995. U1 snRNP-ASF/SF2 interaction and 5' splice site recognition: characterization of required elements. *Nucleic Acids Res*, 23:3260-7.
Johnson, A. W. 1997. Rat1p and Xrn1p are functionally interchangeable exoribonucleases that are restricted to and required in the nucleus and cytoplasm, respectively. *Mol Cell Biol*, 17:6122-30.

Jurica, M. S., and M. J. Moore. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell*, 12:5-14.
Kameoka, S., P. Duque, and M. M. Konarska. 2004. p54(nrb) associates with the 5' splice site within large transcription/splicing complexes. *Embo J*, 23:1782-91.
Kao, H. Y., and P. G. Siliciano. 1996. Identification of Prp40, a novel essential yeast splicing factor associated with the U1 small nuclear ribonucleoprotein particle. *Mol Cell Biol*, 16:960-7.
Kar, A., N. Havlioglu, W. Y. Tarn, and J. Y. Wu. 2006. RBM4 interacts with an intronic element and stimulates tau exon 10 inclusion. *J Biol Chem*, 281:24479-88.
Keast, D. 1968. A simple index for the measurement of the runting syndrome and its use in the study of the influence of the gut flora in its production. *Immunology*, 15:237-45.
Kim, M., N. J. Krogan, L. Vasiljeva, O. J. Rando, E. Nedea, J. F. Greenblatt, and S. Buratowski. 2004. The yeast Rat1 exonuclease promotes transcription termination by RNA polymerase II. *Nature*, 432:517-22.
Kim, S., E. H. Humphries, L. Tjoelker, L. Carlson, and C. B. Thompson. 1990. Ongoing diversification of the rearranged immunoglobulin light-chain gene in a bursal lymphoma cell line. *Mol Cell Biol*, 10:3224-31.
Kim, S. H., and R. J. Lin. 1996. Spliceosome activation by PRP2 ATPase prior to the first transesterification reaction of pre-mRNA splicing. *Mol Cell Biol*, 16:6810-9.
Kim, S. J., D. Gershov, X. Ma, N. Brot, and K. B. Elkon. 2003. Opsonization of apoptotic cells and its effect on macrophage and T cell immune responses. *Ann N Y Acad Sci*, 987:68-78.
Kiss, T. 2004. Biogenesis of small nuclear RNPs. *J Cell Sci*, 117:5949-51.
Kistler, A. L., and C. Guthrie. 2001. Deletion of MUD2, the yeast homolog of U2AF65, can bypass the requirement for sub2, an essential spliceosomal ATPase. *Genes Dev*, 15:42-9.
Kohtz, J. D., S. F. Jamison, C. L. Will, P. Zuo, R. Luhrmann, M. A. Garcia-Blanco, and J. L. Manley. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature*, 368:119-24.
Komarnitsky, P., E. J. Cho, and S. Buratowski. 2000. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev*, 14:2452-60.
Kotovic, K. M., D. Lockshon, L. Boric, and K. M. Neugebauer. 2003. Cotranscriptional recruitment of the U1 snRNP to intron-containing genes in yeast. *Mol Cell Biol*, 23:5768-79.
Kramer, A. 1996. The structure and function of proteins involved in mammalian pre-mRNA splicing. *Annu Rev Biochem*, 65:367-409.
Kramer, A., and U. Utans. 1991. Three protein factors (SF1, SF3 and U2AF) function in pre-splicing complex formation in addition to snRNPs. *Embo J*, 10:1503-9.
Krol, A., H. Gallinaro, E. Lazar, M. Jacob, and C. Branlant. 1981. The nuclear 5S RNAs from chicken, rat and man. U5 RNAs are encoded by multiple genes. *Nucleic Acids Res*, 9:769-87.
Lai, M. C., H. W. Kuo, W. C. Chang, and W. Y. Tarn. 2003. A novel splicing regulator shares a nuclear import pathway with SR proteins. *Embo J*, 22:1359-69.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, et al. 2001. Initial sequencing and analysis of the human genome. *Nature*, 409:860-921.

Lauber, J., P. Fabrizio, S. Teigelkamp, W. S. Lane, E. Hartmann, and R. Luhrmann. 1996. The HeLa 200 kDa U5 snRNP-specific protein and its homologue in Saccharomyces cerevisiae are members of the DEXH-box protein family of putative RNA helicases. *Embo J*, 15:4001-15.

Le Hir, H., M. J. Moore, and L. E. Maquat. 2000. Pre-mRNA splicing alters mRNP composition: evidence for stable association of proteins at exon-exon junctions. *Genes Dev*, 14:1098-108.

Leadbetter, E. A., I. R. Rifkin, A. M. Hohlbaum, B. C. Beaudette, M. J. Shlomchik, and A. Marshak-Rothstein. 2002. Chromatin-IgG complexes activate B cells by dual engagement of IgM and Toll-like receptors. *Nature*, 416:603-7.

Lerner, M. R., J. A. Boyle, J. A. Hardin, and J. A. Steitz. 1981. Two novel classes of small ribonucleoproteins detected by antibodies associated with lupus erythematosus. *Science*, 211:400-2.

Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. A. Steitz. 1980. Are snRNPs involved in splicing? *Nature*, 283:220-4.

Lerner, M. R., and J. A. Steitz. 1979. Antibodies to small nuclear RNAs complexed with proteins are produced by patients with systemic lupus erythematosus. *Proc Natl Acad Sci U S A*, 76:5495-9.

Lewis, J. D., E. Izaurralde, A. Jarmolowski, C. McGuigan, and I. W. Mattaj. 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5' splice site. *Genes Dev*, 10:1683-98.

Libri, D., N. Graziani, C. Saguez, and J. Boulay. 2001. Multiple roles for the yeast SUB2/yUAP56 gene in splicing. *Genes Dev*, 15:36-41.

Lin, C., L. Yang, J. J. Yang, Y. Huang, and Z. R. Liu. 2005. ATPase/helicase activities of p68 RNA helicase are required for pre-mRNA splicing but not for assembly of the spliceosome. *Mol Cell Biol*, 25:7484-93.

Lin, R. J., A. J. Newman, S. C. Cheng, and J. Abelson. 1985. Yeast mRNA splicing in vitro. *J Biol Chem*, 260:14780-92.
Liu, Z., I. Luyten, M. J. Bottomley, A. C. Messias, S. Houngninou-Molango, R. Sprangers, K. Zanier, A. Kramer, and M. Sattler. 2001. Structural basis for recognition of the intron branch site RNA by splicing factor 1. *Science*, 294:1098-102.
Lockhart, S. R., and B. C. Rymond. 1994. Commitment of yeast pre-mRNA to the splicing pathway requires a novel U1 small nuclear ribonucleoprotein polypeptide, Prp39p. *Mol Cell Biol*, 14:3623-33.
Lu, P. J., A. L. Hsu, D. S. Wang, H. Y. Yan, H. L. Yin, and C. S. Chen. 1998. Phosphoinositide 3-kinase in rat liver nuclei. *Biochemistry*, 37:5738-45.
Luhrmann, R., B. Kastner, and M. Bach. 1990. Structure of spliceosomal snRNPs and their role in pre-mRNA splicing. *Biochim Biophys Acta*, 1087:265-92.
Luo, M. J., and R. Reed. 1999. Splicing is required for rapid and efficient mRNA export in metazoans. *Proc Natl Acad Sci U S A*, 96:14937-42.
Luo, M. L., Z. Zhou, K. Magni, C. Christoforides, J. Rappsilber, M. Mann, and R. Reed. 2001. Pre-mRNA splicing and mRNA export linked by direct interactions between UAP56 and Aly. *Nature*, 413:644-7.
Madhani, H. D., and C. Guthrie. 1994. Dynamic RNA-RNA interactions in the spliceosome. *Annu Rev Genet*, 28:1-26.
Makarova, O. V., E. M. Makarov, and R. Luhrmann. 2001. The 65 and 110 kDa SR-related proteins of the U4/U6.U5 tri-snRNP are essential for the assembly of mature spliceosomes. *Embo J*, 20:2553-63.
Mandal, S. S., C. Chu, T. Wada, H. Handa, A. J. Shatkin, and D. Reinberg. 2004. Functional interactions of RNA-capping enzyme with factors that positively and negatively regulate promoter escape by RNA polymerase II. *Proc Natl Acad Sci U S A*, 101:7572-7.
Maniatis, T., and R. Reed. 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature*, 325:673-8.
Maniatis, T., and R. Reed. 2002. An extensive network of coupling among gene expression machines. *Nature*, 416:499-506.
Marrack, P., J. Kappler, and B. L. Kotzin. 2001. Autoimmune disease: why and where it occurs. *Nat Med*, 7:899-905.
Marshak-Rothstein, A. 2006. Toll-like receptors in systemic autoimmune disease. *Nat Rev Immunol*, 6:823-35.
Martin, A., S. Schneider, and B. Schwer. 2002. Prp43 is an essential RNA-dependent ATPase required for release of lariat-intron from the spliceosome. *J Biol Chem*, 277:17743-50.
Martin, D. A., and K. B. Elkon. 2005. Autoantibodies make a U-turn: the toll hypothesis for autoantibody specificity. *J Exp Med*, 202:1465-9.
Mattaj, I. W. 1986. Cap trimethylation of U snRNA is cytoplasmic and dependent on U snRNP protein binding. *Cell*, 46:905-11.
Mattioli, M., and M. Reichlin. 1971. Characterization of a soluble nuclear ribonucleoprotein antigen reactive with SLE sera. *J Immunol*, 107:1281-90.

McCracken, S., N. Fong, K. Yankulov, S. Ballantyne, G. Pan, J. Greenblatt, S. D. Patterson, M. Wickens, and D. L. Bentley. 1997. The C-terminal domain of RNA polymerase II couples mRNA processing to transcription. *Nature*, 385:357-61.

Mitchell, P., and D. Tollervey. 2003. An NMD pathway in yeast involving accelerated deadenylation and exosome-mediated 3'-->5' degradation. *Mol Cell*, 11:1405-13.

Moore, R. E., M. K. Young, and T. D. Lee. 2000. Method for screening peptide fragment ion mass spectra prior to database searching. *J Am Soc Mass Spectrom*, 11:422-6.

Moore, R. E., M. K. Young, and T. D. Lee. 2002. Qscore: an algorithm for evaluating SEQUEST database search results. *J Am Soc Mass Spectrom*, 13:378-86.

Moraitou, M., M. Patrinou-Georgoula, and A. Guialis. 1998. Structural/functional properties of a mammalian multi-component structure containing all major spliceosomal small nuclear ribonucleoprotein particles. *Biochem J*, 332 (Pt 1):135-44.

Moriyama, Y., J. L. Hodnett, A. W. Prestayko, and H. Busch. 1969. Studies on the nuclear 4 to 7 s RNA of the Novikoff hepatoma. *J Mol Biol*, 39:335-49.

Morris, D. P., and A. L. Greenleaf. 2000. The splicing factor, Prp40, binds the phosphorylated carboxyl-terminal domain of RNA polymerase II. *J Biol Chem*, 275:39935-43.

Mouaikel, J., C. Verheggen, E. Bertrand, J. Tazi, and R. Bordonne. 2002. Hypermethylation of the cap structure of both yeast snRNAs and snoRNAs requires a conserved methyltransferase that is localized to the nucleolus. *Mol Cell*, 9:891-901.

Muller, S., B. Wolpensinger, M. Angenitzki, A. Engel, J. Sperling, and R. Sperling. 1998. A supraspliceosome model for large nuclear ribonucleoprotein particles based on mass determinations by scanning transmission electron microscopy. *J Mol Biol*, 283:383-94.

Muller, U. 1999. Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev*, 82:3-21.

Muramatsu, M., J. L. Hodnett, and H. Busch. 1966. Base composition of fractions of nuclear and nucleolar ribonucleic acid obtained by sedimentation and chromatography. *J Biol Chem*, 241:1544-50.

Myer, V. E., X. C. Fan, and J. A. Steitz. 1997. Identification of HuR as a protein implicated in AUUUA-mediated mRNA decay. *Embo J*, 16:2130-9.

Neubauer, G., A. Gottschalk, P. Fabrizio, B. Seraphin, R. Luhrmann, and M. Mann. 1997. Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc Natl Acad Sci U S A*, 94:385-90.

Neubauer, G., A. King, J. Rappsilber, C. Calvio, M. Watson, P. Ajuh, J. Sleeman, A. Lamond, and M. Mann. 1998. Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet*, 20:46-50.

Noble, S. M., and C. Guthrie. 1996. Identification of novel genes required for yeast pre-mRNA splicing by means of cold-sensitive mutations. *Genetics*, 143:67-80.

O'Day, C. L., G. Dalbadie-McFarland, and J. Abelson. 1996. The Saccharomyces cerevisiae Prp5 protein has RNA-dependent ATPase activity with specificity for U2 small nuclear RNA. *J Biol Chem*, 271:33261-7.

O'Shea, J. J., A. Ma, and P. Lipsky. 2002. Cytokines and autoimmunity. *Nat Rev Immunol*, 2:37-45.

Ohi, M. D., A. J. Link, L. Ren, J. L. Jennings, W. H. McDonald, and K. L. Gould. 2002. Proteomics analysis reveals stable multiprotein complexes in both fission and budding yeasts containing Myb-related Cdc5p/Cef1p, novel pre-mRNA splicing factors, and snRNAs. *Mol Cell Biol*, 22:2011-24.

Padgett, R. A., M. M. Konarska, P. J. Grabowski, S. F. Hardy, and P. A. Sharp. 1984. Lariat RNA's as intermediates and products in the splicing of messenger RNA precursors. *Science*, 225:898-903.

Palacios, I., M. Hetzer, S. A. Adam, and I. W. Mattaj. 1997. Nuclear import of U snRNPs requires importin beta. *Embo J*, 16:6783-92.

Palancade, B., and O. Bensaude. 2003. Investigating RNA polymerase II carboxyl-terminal domain (CTD) phosphorylation. *Eur J Biochem*, 270:3859-70.

Paradis, C., P. Cloutier, L. Shkreta, J. Toutant, K. Klarskov, and B. Chabot. 2007. hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. *Rna*.

Park, E., J. Han, G. H. Son, M. S. Lee, S. Chung, S. H. Park, K. Park, K. H. Lee, S. Choi, J. Y. Seong, and K. Kim. 2006. Cooperative actions of Tra2alpha with 9G8 and SRp30c in the RNA splicing of the gonadotropin-releasing hormone gene transcript. *J Biol Chem*, 281:401-9.

Parker, R., P. G. Siliciano, and C. Guthrie. 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, 49:229-39.

Patton, J. G., S. A. Mayer, P. Tempst, and B. Nadal-Ginard. 1991. Characterization and molecular cloning of polypyrimidine tract-binding protein: a component of a complex necessary for pre-mRNA splicing. *Genes Dev*, 5:1237-51.

Paul, C. P., P. D. Good, I. Winer, and D. R. Engelke. 2002. Effective expression of small interfering RNA in human cells. *Nat Biotechnol*, 20:505-8.

Paushkin, S., A. K. Gubitz, S. Massenet, and G. Dreyfuss. 2002. The SMN complex, an assemblyosome of ribonucleoproteins. *Curr Opin Cell Biol*, 14:305-12.

Pease, S., and R. L. Williams. 1990. Formation of germ-line chimeras from embryonic stem cells maintained with recombinant leukemia inhibitory factor. *Exp Cell Res*, 190:209-11.

Petracek, M. E., and M. S. Longtine. 2002. PCR-based engineering of yeast genome. *Methods Enzymol*, 350:445-69.

Pikielny, C. W., and M. Rosbash. 1986. Specific small nuclear RNAs are associated with yeast spliceosomes. *Cell*, 45:869-77.

Plessel, G., U. Fischer, and R. Luhrmann. 1994. m3G cap hypermethylation of U1 small nuclear ribonucleoprotein (snRNP) in vitro: evidence that the U1 small nuclear RNA-(guanosine-N2)-methyltransferase is a non-snRNP cytoplasmic protein that requires a binding site on the Sm core domain. *Mol Cell Biol*, 14:4160-72.

Prestayko, A. W., and H. Busch. 1968. Low molecular weight RNA of the chromatin fraction from Novikoff hepatoma and rat liver nuclei. *Biochim Biophys Acta*, 169:327-37.
Proudfoot, N. 2004. New perspectives on connecting messenger RNA 3' end formation to transcription. *Curr Opin Cell Biol*, 16:272-8.
Puig, O., F. Caspary, G. Rigaut, B. Rutz, E. Bouveret, E. Bragado-Nilsson, M. Wilm, and B. Seraphin. 2001. The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods*, 24:218-29.
Raghunathan, P. L., and C. Guthrie. 1998. RNA unwinding in U4/U6 snRNPs requires ATP hydrolysis and the DEIH-box splicing factor Brr2. *Curr Biol*, 8:847-55.
Raitskin, O., M. Angenitzki, J. Sperling, and R. Sperling. 2002. Large nuclear RNP particles--the nuclear pre-mRNA processing machine. *J Struct Biol*, 140:123-30.
Raitskin, O., D. S. Cho, J. Sperling, K. Nishikura, and R. Sperling. 2001. RNA editing activity is associated with splicing factors in lnRNP particles: The nuclear pre-mRNA processing machinery. *Proc Natl Acad Sci U S A*, 98:6571-6.
Raj, N. B., T. S. Ro-Choi, and H. Busch. 1975. Nuclear ribonucleoprotein complexes containing U1 and U2 RNA. *Biochemistry*, 14:4380-5.
Ramanathan, Y., S. M. Rajpara, S. M. Reza, E. Lees, S. Shuman, M. B. Mathews, and T. Pe'ery. 2001. Three RNA polymerase II carboxyl-terminal domain kinases display distinct substrate preferences. *J Biol Chem*, 276:10913-20.
Randolph, D. A., and C. G. Fathman. 2006. Cd4+Cd25+ regulatory T cells and their therapeutic potential. *Annu Rev Med*, 57:381-402.
Rappsilber, J., U. Ryder, A. I. Lamond, and M. Mann. 2002. Large-scale proteomic analysis of the human spliceosome. *Genome Res*, 12:1231-45.
Reddy, R., D. Henning, and H. Busch. 1979. Nucleotide sequence of nucleolar U3B RNA. *J Biol Chem*, 254:11097-105.
Reddy, R., D. Henning, and H. Busch. 1980. Substitutions, insertions, and deletions in two highly conserved U3 RNA species. *J Biol Chem*, 255:7029-33.
Reddy, R., D. Henning, and H. Busch. 1981a. The primary nucleotide sequence of U4 RNA. *J Biol Chem*, 256:3532-8.
Reddy, R., D. Henning, and H. Busch. 1981b. Pseudouridine residues in the 5'-terminus of uridine-rich nuclear RNA I (U1 RNA). *Biochem Biophys Res Commun*, 98:1076-83.
Reddy, R., T. O. Sitz, T. S. Ro-Choi, and H. Busch. 1974. Two-dimensional polyacrylamide gel electrophoresis separation of low molecular weight nuclear RNA. *Biochem Biophys Res Commun*, 56:1017-22.
Rivkin, E., M. J. Vella, and R. G. Lahita. 1994. A heterogeneous immune response to an SmD-like epitope by SLE patients. *J Autoimmun*, 7:119-32.
Robertson, E., A. Bradley, M. Kuehn, and M. Evans. 1986. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature*, 323:445-8.
Rokeach, L. A., J. A. Haselby, and S. O. Hoch. 1992. Overproduction of a human snRNP-associated Sm-D autoantigen in Escherichia coli and Saccharomyces cerevisiae. *Gene*, 118:247-53.

Rothstein, R. 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. *Methods Enzymol*, 194:281-301.
Rottman, F., A. J. Shatkin, and R. P. Perry. 1974. Sequences containing methylated nucleotides at the 5' termini of messenger RNAs: possible implications for processing. *Cell*, 3:197-9.
Ruby, S. W., and J. Abelson. 1988. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. *Science*, 242:1028-35.
Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green. 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell*, 38:317-31.
Ruskin, B., P. D. Zamore, and M. R. Green. 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. *Cell*, 52:207-19.
Russell, C. S., S. Ben-Yehuda, I. Dix, M. Kupiec, and J. D. Beggs. 2000. Functional analyses of interacting factors involved in both pre-mRNA splicing and cell cycle progression in Saccharomyces cerevisiae. *Rna*, 6:1565-72.
Sakashita, E., S. Tatsumi, D. Werner, H. Endo, and A. Mayeda. 2004. Human RNPS1 and its associated factors: a versatile alternative pre-mRNA splicing regulator in vivo. *Mol Cell Biol*, 24:1174-87.
Sakharkar, M. K., V. T. Chow, and P. Kangueane. 2004. Distributions of exons and introns in the human genome. *In Silico Biol*, 4:387-93.
Schroeder, S. C., B. Schwer, S. Shuman, and D. Bentley. 2000. Dynamic association of capping enzymes with transcribing RNA polymerase II. *Genes Dev*, 14:2435-40.
Schwer, B., and C. H. Gross. 1998. Prp22, a DExH-box RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *Embo J*, 17:2086-94.
Schwer, B., and C. Guthrie. 1991. PRP16 is an RNA-dependent ATPase that interacts transiently with the spliceosome. *Nature*, 349:494-9.
Schwer, B., and C. Guthrie. 1992. A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *Embo J*, 11:5033-9.
Seaman, F., P. Sawhney, C. J. Giammona, and J. H. Richburg. 2003. Cisplatin-induced pulse of germ cell apoptosis precedes long-term elevated apoptotic rates in C57/BL/6 mouse testis. *Apoptosis*, 8:101-8.
Sekeris, C. E., and J. Niessing. 1975. Evidence for the existence of a structural RNA component in the nuclear ribonucleoprotein particles containing heterogeneous RNA. *Biochem Biophys Res Commun*, 62:642-50.
Seraphin, B., L. Kretzner, and M. Rosbash. 1988. A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5' cleavage site. *Embo J*, 7:2533-8.
Sharp, P. A. 1994. Split genes and RNA splicing. *Cell*, 77:805-15.
Shatkin, A. J. 1976. Capping of eucaryotic mRNAs. *Cell*, 9:645-53.
Shibata, H., T. S. Ro-Choi, R. Reddy, Y. C. Choi, D. Henning, and H. Busch. 1975. The primary nucleotide sequence of nuclear U-2 ribonucleic acid. The 5'-terminal portion of the molecule. *J Biol Chem*, 250:3909-20.
Shichijo, S., M. Nakao, Y. Imai, H. Takasu, M. Kawamoto, F. Niiya, D. Yang, Y. Toh, H. Yamana, and K. Itoh. 1998. A gene encoding antigenic peptides of human

squamous cell carcinoma recognized by cytotoxic T lymphocytes. *J Exp Med*, 187:277-88.
Singh, R., and R. Reddy. 1989. Gamma-monomethyl phosphate: a cap structure in spliceosomal U6 small nuclear RNA. *Proc Natl Acad Sci U S A*, 86:8280-3.
Sperling, R., J. Sperling, A. D. Levine, P. Spann, G. R. Stark, and R. D. Kornberg. 1985. Abundant nuclear ribonucleoprotein form of CAD RNA. *Mol Cell Biol*, 5:569-75.
Staikou, E. V., J. G. Routsias, A. A. Makri, A. Terzoglou, M. Sakarellos-Daitsiotis, C. Sakarellos, G. Panayotou, H. M. Moutsopoulos, and A. G. Tzioufas. 2003. Calreticulin binds preferentially with B cell linear epitopes of Ro60 kD autoantigen, enhancing recognition by anti-Ro60 kD autoantibodies. *Clin Exp Immunol*, 134:143-50.
Staley, J. P., and C. Guthrie. 1998. Mechanical devices of the spliceosome: motors, clocks, springs, and things. *Cell*, 92:315-26.
Staley, J. P., and C. Guthrie. 1999. An RNA switch at the 5' splice site requires ATP and the DEAD box protein Prp28p. *Mol Cell*, 3:55-64.
Stephens, R. M., and T. D. Schneider. 1992. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J Mol Biol*, 228:1124-36.
Stevens, S. W. 2000. Analysis of low-abundance ribonucleoprotein particles from yeast by affinity chromatography and mass spectrometry microsequencing. *Methods Enzymol*, 318:385-98.
Stevens, S. W., and J. Abelson. 1999. Purification of the yeast U4/U6.U5 small nuclear ribonucleoprotein particle and identification of its proteins. *Proc Natl Acad Sci U S A*, 96:7226-31.
Stevens, S. W., I. Barta, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee, and J. Abelson. 2001. Biochemical and genetic analyses of the U5, U6, and U4/U6 x U5 small nuclear ribonucleoproteins from Saccharomyces cerevisiae. *Rna*, 7:1543-53.
Stevens, S. W., D. E. Ryan, H. Y. Ge, R. E. Moore, M. K. Young, T. D. Lee, and J. Abelson. 2002. Composition and functional characterization of the yeast spliceosomal penta-snRNP. *Mol Cell*, 9:31-44.
Strasser, K., and E. Hurt. 2001. Splicing factor Sub2p is required for nuclear mRNA export through its interaction with Yra1p. *Nature*, 413:648-52.
Strasser, K., S. Masuda, P. Mason, J. Pfannstiel, M. Oppizzi, S. Rodriguez-Navarro, A. G. Rondon, A. Aguilera, K. Struhl, R. Reed, and E. Hurt. 2002. TREX is a conserved complex coupling transcription with messenger RNA export. *Nature*, 417:304-8.
Strauss, E. J., and C. Guthrie. 1991. A cold-sensitive mRNA splicing mutant is a member of the RNA helicase gene family. *Genes Dev*, 5:629-41.
Strauss, E. J., and C. Guthrie. 1994. PRP28, a 'DEAD-box' protein, is required for the first step of mRNA splicing in vitro. *Nucleic Acids Res*, 22:3187-93.
Takahashi, T., and S. Sakaguchi. 2003. The role of regulatory T cells in controlling immunologic self-tolerance. *Int Rev Cytol*, 225:1-32.
Tan, E. M., and H. G. Kunkel. 1966. Characteristics of a soluble nuclear antigen precipitating with sera of patients with systemic lupus erythematosus. *J Immunol*, 96:464-71.

Teigelkamp, S., T. Achsel, C. Mundt, S. F. Gothel, U. Cronshagen, W. S. Lane, M. Marahiel, and R. Luhrmann. 1998. The 20kD protein of human [U4/U6.U5] tri-snRNPs is a novel cyclophilin that forms a complex with the U4/U6-specific 60kD and 90kD proteins. *Rna*, 4:127-41.

Teigelkamp, S., C. Mundt, T. Achsel, C. L. Will, and R. Luhrmann. 1997. The human U5 snRNP-specific 100-kD protein is an RS domain-containing, putative RNA helicase with significant homology to the yeast splicing factor Prp28p. *Rna*, 3:1313-26.

Terzoglou, A. G., J. G. Routsias, S. Avrameas, H. M. Moutsopoulos, and A. G. Tzioufas. 2006a. Preferential recognition of the phosphorylated major linear B-cell epitope of La/SSB 349-368 aa by anti-La/SSB autoantibodies from patients with systemic autoimmune diseases. *Clin Exp Immunol*, 144:432-9.

Terzoglou, A. G., J. G. Routsias, H. M. Moutsopoulos, and A. G. Tzioufas. 2006b. Post-translational modifications of the major linear epitope 169-190aa of Ro60 kDa autoantigen alter the autoantibody binding. *Clin Exp Immunol*, 146:60-5.

Thomas, K. R., and M. R. Capecchi. 1987. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*, 51:503-12.

Thompson, C. B., E. H. Humphries, L. M. Carlson, C. L. Chen, and P. E. Neiman. 1987. The effect of alterations in myc gene expression on B cell development in the bursa of Fabricius. *Cell*, 51:371-81.

Torgerson, T. R. 2006. Regulatory T cells in human autoimmune diseases. *Springer Semin Immunopathol*, 28:63-76.

Tsai, W. Y., Y. T. Chow, H. R. Chen, K. T. Huang, R. I. Hong, S. P. Jan, N. Y. Kuo, T. Y. Tsao, C. H. Chen, and S. C. Cheng. 1999. Cef1p is a component of the Prp19p-associated complex and essential for pre-mRNA splicing. *J Biol Chem*, 274:9455-62.

Utans, U., S. E. Behrens, R. Luhrmann, R. Kole, and A. Kramer. 1992. A splicing factor that is inactivated during in vivo heat shock is functionally equivalent to the [U4/U6.U5] triple snRNP-specific proteins. *Genes Dev*, 6:631-41.

Utz, P. J., T. J. Gensler, and P. Anderson. 2000. Death, autoantigen modifications, and tolerance. *Arthritis Res*, 2:101-14.

Utz, P. J., M. Hottelet, P. H. Schur, and P. Anderson. 1997. Proteins phosphorylated during stress-induced apoptosis are common targets for autoantibody production in patients with systemic lupus erythematosus. *J Exp Med*, 185:843-54.

Vagner, S., C. Vagner, and I. W. Mattaj. 2000. The carboxyl terminus of vertebrate poly(A) polymerase interacts with U2AF 65 to couple 3'-end processing and splicing. *Genes Dev*, 14:403-13.

Valenta, R., S. Natter, S. Seiberler, S. Wichlas, D. Maurer, M. Hess, M. Pavelka, M. Grote, F. Ferreira, Z. Szepfalusi, P. Valent, and G. Stingl. 1998. Molecular characterization of an autoallergen, Hom s 1, identified by serum IgE from atopic dermatitis patients. *J Invest Dermatol*, 111:1178-83.

Vijayraghavan, U., M. Company, and J. Abelson. 1989. Isolation and characterization of pre-mRNA splicing mutants of Saccharomyces cerevisiae. *Genes Dev*, 3:1206-16.

Vollmer, J., S. Tluk, C. Schmitz, S. Hamm, M. Jurk, A. Forsbach, S. Akira, K. M. Kelly, W. H. Reeves, S. Bauer, and A. M. Krieg. 2005. Immune stimulation mediated by

autoantigen binding sites within small nuclear RNAs involves Toll-like receptors 7 and 8. *J Exp Med*, 202:1575-85.
von Muhlen, C. A., and E. M. Tan. 1995. Autoantibodies in the diagnosis of systemic rheumatic diseases. *Semin Arthritis Rheum*, 24:323-58.
Wagner, J. D., E. Jankowsky, M. Company, A. M. Pyle, and J. N. Abelson. 1998. The DEAH-box protein PRP22 is an ATPase that mediates ATP-dependent mRNA release from the spliceosome and unwinds RNA duplexes. *Embo J*, 17:2926-37.
Wang, P., P. J. Lou, S. Leu, and P. Ouyang. 2002. Modulation of alternative pre-mRNA splicing in vivo by pinin. *Biochem Biophys Res Commun*, 294:448-55.
Wassarman, D. A., and J. A. Steitz. 1993. A base-pairing interaction between U2 and U6 small nuclear RNAs occurs in > 150S complexes in HeLa cell extracts: implications for the spliceosome assembly pathway. *Proc Natl Acad Sci U S A*, 90:7139-43.
Wellmann, U., A. Werner, and T. H. Winkler. 2001. Altered selection processes of B lymphocytes in autoimmune NZB/W mice, despite intact central tolerance against DNA. *Eur J Immunol*, 31:2800-10.
White, P. J., W. D. Gardner, and S. O. Hoch. 1981. Identification of the immunogenically active components of the Sm and RNP antigens. *Proc Natl Acad Sci U S A*, 78:626-30.
Will, C. L., and R. Luhrmann. 2001. Spliceosomal UsnRNP biogenesis, structure and function. *Curr Opin Cell Biol*, 13:290-301.
Will, C. L., C. Schneider, M. Hossbach, H. Urlaub, R. Rauhut, S. Elbashir, T. Tuschl, and R. Luhrmann. 2004. The human 18S U11/U12 snRNP contains a set of novel proteins not found in the U2-dependent spliceosome. *Rna*, 10:929-41.
Will, C. L., H. Urlaub, T. Achsel, M. Gentzel, M. Wilm, and R. Luhrmann. 2002. Characterization of novel SF3b and 17S U2 snRNP proteins, including a human Prp5p homologue and an SF3b DEAD-box protein. *Embo J*, 21:4978-88.
Williams, D. G., N. G. Sharpe, G. Wallace, and D. S. Latchman. 1990. A repeated proline-rich sequence in Sm B/B' and N is a dominant epitope recognized by human and murine autoantibodies. *J Autoimmun*, 3:715-25.
Winding, P., and M. W. Berchtold. 2001. The chicken B cell line DT40: a novel tool for gene disruption experiments. *J Immunol Methods*, 249:1-16.
Wither, J. E., V. Roy, and L. A. Brennan. 2000. Activated B cells express increased levels of costimulatory molecules in young autoimmune NZB and (NZB x NZW)F(1) mice. *Clin Immunol*, 94:51-63.
Wolin, S. L., and J. A. Steitz. 1984. The Ro small cytoplasmic ribonucleoproteins: identification of the antigenic protein and its binding site on the Ro RNAs. *Proc Natl Acad Sci U S A*, 81:1996-2000.
Xu, Q., B. Modrek, and C. Lee. 2002. Genome-wide detection of tissue-specific alternative splicing in the human transcriptome. *Nucleic Acids Res*, 30:3754-66.
Xu, Y. Z., C. M. Newnham, S. Kameoka, T. Huang, M. M. Konarska, and C. C. Query. 2004. Prp5 bridges U1 and U2 snRNPs and enables stable U2 snRNP association with intron RNA. *Embo J*, 23:376-85.

Yan, D., R. Perriman, H. Igel, K. J. Howe, M. Neville, and M. Ares, Jr. 1998. CUS2, a yeast homolog of human Tat-SF1, rescues function of misfolded U2 through an unusual RNA recognition motif. *Mol Cell Biol*, 18:5000-9.

Yang, V. W., M. R. Lerner, J. A. Steitz, and S. J. Flint. 1981. A small nuclear ribonucleoprotein is required for splicing of adenoviral early RNA sequences. *Proc Natl Acad Sci U S A*, 78:1371-5.

Yitzhaki, S., E. Miriami, R. Sperling, and J. Sperling. 1996. Phosphorylated Ser/Arg-rich proteins: limiting factors in the assembly of 200S large nuclear ribonucleoprotein particles. *Proc Natl Acad Sci U S A*, 93:8830-5.

Zahler, A. M., and M. B. Roth. 1995. Distinct functions of SR proteins in recruitment of U1 small nuclear ribonucleoprotein to alternative 5' splice sites. *Proc Natl Acad Sci U S A*, 92:2642-6.

Zamore, P. D., and M. R. Green. 1989. Identification, purification, and biochemical characterization of U2 small nuclear ribonucleoprotein auxiliary factor. *Proc Natl Acad Sci U S A*, 86:9243-7.

Zeng, D., M. K. Lee, J. Tung, A. Brendolan, and S. Strober. 2000. Cutting edge: a role for CD1 in the pathogenesis of lupus in NZB/NZW mice. *J Immunol*, 164:5000-4.

Zhang, D., and M. Rosbash. 1999. Identification of eight proteins that cross-link to pre-mRNA in the yeast commitment complex. *Genes Dev*, 13:581-92.

Zhang, M., and M. R. Green. 2001. Identification and characterization of yUAP/Sub2p, a yeast homolog of the essential human pre-mRNA splicing factor hUAP56. *Genes Dev*, 15:30-5.

Zhou, Z., L. J. Licklider, S. P. Gygi, and R. Reed. 2002. Comprehensive proteomic analysis of the human spliceosome. *Nature*, 419:182-5.

Zhou, Z., M. J. Luo, K. Straesser, J. Katahira, E. Hurt, and R. Reed. 2000. The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature*, 407:401-5.

Zhuang, Y., and A. M. Weiner. 1986. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell*, 46:827-35.

Zhuang, Y., and A. M. Weiner. 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev*, 3:1545-52.

Zhuang, Y. A., A. M. Goldstein, and A. M. Weiner. 1989. UACUAAC is the preferred branch site for mammalian mRNA splicing. *Proc Natl Acad Sci U S A*, 86:2752-6.

Vita

Yen-I Grace Chen was born in Chia-Yi, Taiwan on September 19, 1977, the daughter of Horng-Shi Chen and Li-Hua Shih. After graduating from First Girl's Highs school in 1995, she attended Taipei Medical University, Taiwan, where she received the degree of Doctor of Medicine in June 2002. In August 2002, she entered the Graduate School at the University of Texas at Austin.

Permanent address: 12F-2, 82, Sec 2, Fuhsing S Rd. Taipei, Taiwan, 106

This dissertation was typed by the author.